ANTIGEN RECEPTORS ON LYMPHOCYTES

Thesis by

GERALD SIU

In Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

Division of Chemistry and Chemical Engineering

California Institute of Technology

Pasadena, California

1986

(submitted 2 December, 1985)

To B.L. and J.L., for being what

I fear I may never be

To see a world in a grain of sand

And a heaven in a wild flower

Hold infinity in the palm of your hand

And eternity in an hour.

William Blake

Fool: The reason why the seven stars are no more than seven is a pretty reason.

King Lear: Because they are not eight?

Fool: Yes, indeed: thou wouldst make a good fool.

King Lear, Act II, scene V

This is a system!?

Mitch Kronenberg

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS

According to the thesis instructions, this section is primarily for acknowledging funding support and assistance in the laboratory. Customarily, this section is fairly brief. However, I have so many people that I wish to thank for making my five years here so enjoyable that I must risk offending anyone who cares to read this thesis and thank all of them. Besides, I have never been able to write anything short in my life.

I wish to thank, of course, my research advisor and friend Lee Hood, whose advice, despite his dislike of the sacred semicolon and his appalling penchant for exercise, proved essential to me; (there you go, Lee, another semicolon) I owe him a lot for helping me learn about science and immunology. I would also like to thank my committee: Norm Davidson, Peter Dervan, and Elliot Meyerowitz for having patience with a befuddled student. The Man from Texas, Steve Crews, taught me about immunoglobulins, sequences, patience, barbeque and countless other things; thanks, Steve, for taking the time and the considerable effort. I would also like to acknowledge four friends of mine whose certifiable insanity could only have been permitted in science: Rick Barth, Joan Kobori, Mitchell Kronenberg, and Paul "Jochie" Tempst. Rick's incredible (but verifiable) stories, Joanie's cat stories (aargh), Mitch's loony jokes that could always be guaranteed to crack me up, and Paul's informative lessons in jazz and flemish will always remain in my mind when I try to think about science. I thank my tolerant coworkers: my ever-patient immunoglobulin collaborator Liz Springer, Steve, the dependable Erich Strauss (and I'll always be sorry for taking away your vacations, although perhaps not nearly as sorry as I should be), Mitchell, and Tak Mak, without whom this thesis would have been impossible. Marilyn Seibel Tomich, Carol Graham Parker, and Dr. Suzanna Horvath provided many oligonucleotides and Debbie Maloney and Karyl Minard provided sequencing gels, both of which saved my

posterier end countless times. I thank my roommates, David "Sticks" Beratan, Michael King, and Markus Meister for putting up with my strange habits and my even stranger tastes in music. I have had the great privilege to be friends with a large number of people both within and without Caltech; I thank them all for a great time and keeping me relatively sane for the last five years. My sincere apologies to the people in the offices of Church and Braun for my unconventional behavior and my off-key renditions of 'O Canada,' and to Connie Katz and Bertha Jones especially, for my abuse of them during the course of this thesis. Many thanks are due to Bertha, Jessie Walker, and Ria de Bruyn for making the Hood lab work, and making Caltech a nice place to go after you've been kicked out of bed in the morning.

Finally, I would like to thank my family: my brothers Larry and Steve, and my parents, Lydia and Patrick, for being supportive and patient, always patient.

ABSTRACT

The structure and evolution of a small V_H gene family called the T15 family was analyzed. It was determined that although selection pressure appeared to be operating to maintain the coding region sequence of these V_H gene segments, these gene segments were diverging from one another very rapidly. Sequences were identified in the 5' flanking region that were conserved between all V_H gene segments and were hypothesized to be important for immunoglobulin heavy-chain gene transcription. Related sequences were identified in immunoglobulin V_L gene segment and histone H2B 5' flanking regions, implying coordinate expression between these genes and the immunoglobulin heavy chain genes.

The structure, organization, evolution, and the generation of diversity in the genes encoding the T-cell antigen receptor were analyzed. The T-cell antigen receptor consists of two chains, referred to as the α and β chains. Each chain consists of two regions, a variable region and a constant region, that are encoded by two different genes. The gene that encodes the variable region of the β chain was found to consist of three gene segments, denoted V_β, D_β, and J_β. The V_β gene segment encodes the first 280-300 bp, the D_β gene segment encodes the next 10-15 bp, and the J_β gene segment encodes the final 50 bp of the variable region gene. The rearrangement event that juxtaposes these gene segments during lymphocyte differentiation appears to be mediated by the same recognition signals that mediate immunoglobulin V gene rearrangement.

Diversity was found to be generated in at least three different manners in the V_β gene. Combinatorial joining permits the rearrangement of different V, D and J gene segments to each other to provide different V gene sequences. Deletion of nucleotides from the ends of the germline gene segments and the random addition of nucleotides at the junction of the rearrangement event are two other mechanisms for generating diversity. A comparison of a rearranged V gene

with the corresponding germline gene segments showed that with the exception of the junctions, the sequences were identical. Therefore, there is no evidence that somatic hypermutation, the random addition of point mutations to the V gene during late stages of B lymphocyte differentiation, is utilized by the T-cell antigen receptor as it is by immunoglobulins.

The initial stage of V_β gene formation was found to be the rearrangement of the D_β gene segment to the J_β gene segment. Both germline D_β gene segments appear to have promoters in the 5' flanking regions that can often result in the transcription of a 1.0 kb mRNA containing D_β-J_β-C_β sequences after D_β-J_β rearrangement. This 1.0 kb mRNA message is present at a high level in the thymus but at lower levels in the spleen, lymph nodes, and in mature T cells, implying that this message or a protein product encoded by this message may be important in T cell ontogeny.

Analysis of the protein sequences of the variable regions of the α and β chains revealed conserved amino acids that are found in all variable region genes. Many of these amino acids were found to be important for V domain structure in immunoglobulins and may be important for the structure of the V_α-V_β domain as well. In addition, analyses of the β-strand forming potential and the relative hydrophobicity of the side chains of the amino acids that make up the T-cell antigen receptor variable regions have indicated that these properties are very similar to those of the immunoglobulin variable regions. These analyses indicate that the immunoglobulin and T-cell receptor antigen-binding regions may be very similar in structure to each other.

CHAPTER ONE

INTRODUCTION

The immune system

The immune response is the primary mechanism utilized in the defense against invading organisms, the destruction of neoplastic cells, and the rejection of foreign grafts. This response is mediated by two distinct populations of cells: B lymphocytes and T lymphocytes. B lymphocytes, responsible for the release of antibodies, originate and mature in the bone marrow and accumulate in the spleen (reviewed in Whitlock *et al.*, 1985). T lymphocytes can be subdivided into three classes: T helper cells (T_H) respond to antigen by releasing factors that stimulate a B cell specific for the same antigen to release antibody and to proliferate (Mitchell and Miller, 1968; Cantor and Boyse, 1975), T suppressor cells (T_S) release factors that inhibit these B cell processes (Gershon, 1974), and cytotoxic T cells (T_C) recognize and kill virally-infected cells and reject foreign grafts (Cerottini *et al.*, 1979). T lymphocytes originate in the bone marrow and subsequently migrate to the thymus. The vast majority of the T cells die in the thymus; the small fraction of survivors become immunocompetent and migrate to the spleen and other peripheral lymphatic organs (reviewed in Scollay *et al.*, 1984; Rothenberg and Lugo, 1985).

Both B and T lymphocytes are activated upon recognition of antigen by their cell-surface antigen receptor. In the case of the B cell, the receptor is the membrane-bound immunoglobulin; in the case of the T cell, the receptor is the antigen-specific T-cell receptor. Although the immune system is capable of recognizing and responding to a seemingly infinite number of antigens, individual lymphocytes can recognize only a limited number of antigenic determinants. The wide diversity of the immune response as a whole results from a large population of different lymphocytes, each capable of recognizing a different set of determinants. The nature of the specificity of lymphocytes and the mechanisms for generating the diversity necessary for the complete immune response has been studied for many years.

Analysis of immunoglobulin proteins

The initial biochemical work on lymphocyte diversity was conducted almost solely on immunoglobulins. Structural analyses of antibody polypeptides defined the existence of two chains that make up the immunoglobulin: the light chain, and the heavy chain (Figure 1). Using different antisera that bound immunoglobulin, it was possible to subdivide the light chains into two classes, κ and λ, and to subdivide the heavy chain into eight classes in mice ($Ig\mu$, $Ig\delta$, $Ig\gamma 1$, $Ig\gamma 2A$, $Ig\gamma 2B$, $Ig\gamma 3$, $Ig\varepsilon$, and $Ig\alpha$) and nine classes in humans ($Ig\mu$, $Ig\delta$, $Ig\gamma 1$, $Ig\gamma 2$, $Ig\gamma 3$, $Ig\gamma 4$, $Ig\varepsilon$, $Ig\alpha 1$ and $Ig\alpha 2$) (for review, see Lennox and Cohn, 1967). Sequence studies of λ chains that identified variable and constant regions (Hilschmann and Craig, 1965; for review, see Putnam *et al.*, 1971) led to the revolutionary proposal that the immunoglobulin chains are encoded by two genes, one for the variable portion and the other for the constant portion (Dreyer and Bennett, 1965). More detailed structural analysis revealed that the amino-terminal variable regions of both heavy and light chains contain three short regions of high variability (Wu and Kabat, 1970; Capra and Kehoe, 1974) which were referred to as hypervariable regions (Figure 1). Analyses of the three-dimensional structure of the immunoglobulin molecule indicated that these hypervariable regions comprised the antigen-contact regions (Padlan *et al.*, 1973; Amzel *et al.*, 1974; Segal *et al.*, 1974). The remaining portions of the variable region were found to be relatively less variable and were referred to as the framework regions (Figure 1). Sequence comparisons of the different variable regions from the heavy, lambda, or kappa chain families revealed that each consisted of different subgroups of V regions that had similar amino acid sequences and sequence alignments (Hood *et al.*, 1968; Milstein, 1967; Cohn *et al.*, 1974). While the different members of each subgroup were homologous, V regions from different subgroups showed only slight homology to each other. It was also observed that every individual was capable of producing

immunoglobulins that utilized the same set of subgroups (Grant and Hood, 1971). These data indicated that a large number of germline genes existed to encode the many different variable regions. Conversely, a catalogue of the differences between kappa variable regions suggested that each appeared to be random; this implied that a limited number of V region genes existed, and that somatic mechanisms generated the V region diversity (Lennox and Cohn, 1967; Smithies, 1967; Hood *et al.*, 1976). Sequence analyses of a large number of heavy and light chain variable regions from different myelomas that bound phosphorylcholine suggested that all utilized closely related, but distinct V regions, and that the variability between the different V regions appeared to be correlated to the type of constant region utilized by the immunoglobulin (Gearhart *et al.*, 1981). Thus, there was strong evidence to indicate the existence of both a large number of germline genes and a somatic mutational mechanism to generate immunoglobulin diversity.

Genomic organization of the immunoglobulin genes

With the development of recombinant DNA techniques, it became possible to study the problem of antibody diversity on the nucleic acid level. There are two separate light chain gene families: kappa, and lambda. The light chain is encoded by two genes, denoted V_L and C_L, that encode the variable region and the constant region, respectively (Hozumi and Tonegawa, 1976). The gene that encodes the variable region actually consists of two gene segments, denoted V_L and J_L, that are separate in the genome and are brought together by DNA rearrangement to form the complete V_L gene during lymphocyte differentiation (Figure 2) (Brack *et al.*, 1978, Seidman *et al.*, 1979, Sakano *et al.*, 1980). There are 90-320 murine V_κ gene segments and perhaps only 15-20 human V_κ gene segments (Bentley and Rabbitts, 1980; Cory *et al.*, 1981; Zeelon *et al.*, 1981).

These V gene segments can be subdivided into subfamilies of gene segments that share 75% or greater homology on the DNA level; in the mouse, there are at least five subfamilies, ranging in size from two to greater than 20 members (Cory *et al.*, 1981). There are four functional J_κ gene segments and one pseudo J_κ gene segment in the mouse genome, and five functional human J_κ gene segments (Figure 3) (Max *et al.*, 1981; Hieter *et al.*, 1982). There is only one kappa constant region gene in both mice and humans, consisting of a single exon (Altenburger *et al.*, 1981; Max *et al.*, 1981). The J_κ gene segments are clustered 2.6 kb 5' to the C_κ gene, with the V_κ gene segments located an undetermined distance 5' to the J_κ cluster. There appear to be two V_λ gene segments, three functional and two pseudo J_λ gene segments, and four C_λ genes in the mouse genome (Bernard *et al.*, 1978; Brack *et al.*, 1978; Blomberg *et al.*, 1981; Blomberg and Tonegawa, 1982; Miller *et al.*, 1982; Tonegawa *et al.*, 1978; Wu *et al.*, 1982). The organization of the lambda locus is more complicated than the kappa locus; the four constant region genes each have a single functional J_λ gene segment in the immediate 5' flanking region, with the exception of $C_{\lambda 3}$, which has one functional J_λ and one pseudo J_λ gene segment. Analysis of many lambda chains indicated that one V_λ gene segment always associated with two of the C_λ genes, while the other V_λ gene segment always associated with the other two C_λ genes (Appella, 1971; Weigert and Riblet, 1976; Dugan *et al.*, 1973; Elliot *et al.*, 1981); these data led to the hypothesis that the lambda locus was organized as V-J-C-J-C-V-J-C-J-C (see Figure 3) (Elliot *et al.*, 1982).

The genes that encode the heavy chain are structurally very similar to the light chain genes. The heavy chain variable region is encoded by three gene segments, V_H, D_H, and J_H, that rearrange to form the complete V_H gene (Figure 2) (Early *et al.*, 1980; Sakano *et al.*, 1980). There are a large number of V_H gene segments; recent studies using C_oT analyses indicate that the V_H gene

family has over 1000 members (D. Livant *et al.*, in preparation). Like the V_λ genes, the V_H gene family can be subdivided into at least seven subfamilies ranging in size from four to 1000 members (Crews *et al.*, 1981; Brodeur and Riblet, 1984; Dildrop, 1984; Livant *et al.*, in preparation). Gene segments within a subfamily appear to be clustered; the average spacing distance appears to be approximately 23 kb, although it can range from 2 kb to greater than 60 kb (S. Crews, E. Springer, G. Siu, and L. Hood, unpublished). There are perhaps 11 D_H gene segments subdivided into at least three different subfamilies in mice (Kurosawa and Tonegawa, 1982), and four D_H gene segments in two subfamilies in humans (Siebenlist *et al.*, 1981). There are four mouse J_H gene segments, and six functional and three nonfunctional human J_H gene segments (Bernard and Gough, 1980; Early *et al.*, 1980; Sakano *et al.*, 1980; Gough and Bernard, 1981; Ravetch *et al.*, 1981), located approximately 6 kb 5' to the C_μ genes. The first D_H gene segment, D_{Q52}, is located 700 bp 5' to J_{H1} in mice, and between $J_{\psi 1}$ and J_{H1} in humans (Sakano *et al.*, 1981; Ravetch *et al.*, 1981). There are eight murine constant region genes, μ, δ, $\gamma 3$, $\gamma 1$, $\gamma 2B$, $\gamma 2a$, ε, and α (Shimizu *et al.*, 1982), each corresponding to a different antibody class (Figure 3). The human constant region locus consist of nine functional genes, μ, δ, $\gamma 1$, $\gamma 2$, $\gamma 3$, $\gamma 4$, ε, $\alpha 1$, and $\alpha 2$ (Ellison et al, 1981, Ellison and Hood, 1982; Rabbitts *et al.*, 1981; Ellison *et al.*, 1982; Takahashi *et al.*, 1982) and three pseudogenes: one homologous to C_γ, and two homologous to C_ε (Battey *et al.*, 1982; Flanagan *et al.*, 1982; Max *et al.*, 1982; Nishida *et al.*, 1982; Takahashi *et al.*, 1982; Ueda *et al.*, 1982). The C_H genes are composed of either three (δ, α) or four ($\gamma 1$, $\gamma 2A$, $\gamma 2B$, $\gamma 3$, ε, and μ) exons, each encoding a functional and structural unit of the heavy chain (Honjo *et al.*, 1979; Tucker *et al.*, 1979; Yamawaki-Katoaka *et al.*, 1980; Ollo *et al.*, 1981; Tucker *et al.*, 1981; Yamawaki-Kataoka *et al.*, 1981; Cheng *et al.*, 1982; Ellison *et al.*, 1982; Ellison and Hood, 1982; Ishida *et al.*, 1982). The difference between the secreted

immunoglobulin and the membrane-bound immunoglobulin is encoded by the last exon of the constant region gene; there are two different 3' terminal sequences encoded in the genome, one that specifies an amino-acid sequence appropriate for membrane-binding, and the other for secretion. The secreted exon is located directly adjacent to the penultimate constant region exon, and the membrane exon is located downstream from the secreted exon. The membrane form of immunoglobulin is formed by two additional RNA splicing events that juxtapose the membrane exon with the remaining C_H exons (Figure 2) (Alt *et al.*, 1980; Early *et al.*, 1980b; Rogers *et al.*, 1980).

During the differentiation of a single B cell, the functional rearranged V_H gene is first expressed with the C_μ gene of the same allelic chromosome. Later in the development of the same lymphocyte, the same V_H gene is often expressed with a different C_H gene (for review, see Davis *et al.*, 1980b). Analysis of germline and rearranged heavy-chain genes revealed that the intervening DNA between the 5' flanking region of C_μ and the 5' flanking region of the newly expressed C_H gene is deleted. This led to the proposal that the complete heavy-chain gene is formed by two recombination events. The first is the V-D-J rearrangement that forms the variable region. The second, called the class switch, rearranges a fully assembled V_H-D_H-J_H gene from one constant region to another (Figure 4). The class switch appears to be mediated by repeat sequences in the 5' flanking region of the C_H genes (Davis *et al.*, 1980a; Kataoka *et al.*, 1980; Maki *et al.*, 1980).

Analysis of the flanking regions of the germline heavy chain and light chain gene segments revealed short sequences in the 3' flanking region of the V gene segments, the 5' flanking region of the J gene segments, and for heavy chains, both 5' and 3' flanking regions of the D gene segments that are highly conserved (Figure 5). These sequences consist of a conserved 7bp sequence, CACAGTG or

its complement CACTGTG, located directly flanking the gene segment, and a relatively conserved 9 bp sequence, CACAAACCC or GGTTTTTGT, located distal to the gene segment. The presence of these sequences in the flanking regions of all of the germline immunoglobulin gene segments led to the proposal that these sequences served as recognition signals for the rearrangement process for V gene formation (Max *et al.*, 1979; Sakano *et al.*, 1979; Early *et al.*, 1980; Sakano *et al.*, 1980). The distance between the 7 bp sequence and the 9 bp sequence is unconserved in sequence, but is highly conserved in length. The spacer sequence length can either be 12 ± 1 bp or 23 ± 1 bp; rearrangement occurs only between a gene segment that has a 12 bp spacer in its recognition signal, and a gene segment that has a 23 bp spacer in its recognition signal (Early *et al.*, 1980; Sakano *et al.*, 1980). It was observed that 12 bp is approximately one complete turn of the DNA helix, and 23 bp is approximately two turns of the DNA helix; these two signals are therefore located on the same side of the DNA helix as if they formed a continuous stretch of 16 conserved nucleotides. Because of the conservation of length of the spacing distances, the rule for proper rearrangement is referred to as the "one turn-two turn" rule (Early *et al.*, 1980). The V_κ gene segments have one-turn signals, and the J_κ gene segments two-turn signals; the V_λ gene segments have two-turn signals, and the J_λ gene segments one-turn signals; the V_H and J_H gene segments have two-turn signals, and the D_H gene segments have one turn signals in both flanking regions (Figure 5) (Early *et al.*, 1980; Sakano *et al.*, 1980; Kurosawa and Tonegawa, 1982).

Mechanisms for generating diversity in immunoglobulin genes

There are six major mechanisms for generating diversity in the immunoglobulin variable region.

1. Germline diversity: There are a large number of germline V, D and J gene segments that can be used to make up a variable region gene.

2. Combinatorial joining: The different V gene segments are capable of associating with any D or J gene segment (Schilling *et al.*, 1980; Weigert *et al.*, 1980). If it is assumed that there are 1000 V_H gene segments, 10 D_H gene segments, and four J_H gene segments, there would be a total of 1000 x 10 x 4 or 40,000 V_H genes possible. If it is assumed that there are 250 V_κ gene segments, and four functional J_κ gene segments, then approximately 1000 V_κ genes would be possible.

3. Combinatorial association: If any light chain could associate with any heavy chain, then there are a total of 40,000 x 1000 or 4×10^7 antibodies possible. Germline diversity, combinatorial joining and combinatorial association, therefore, can contribute significantly to immunoglobulin diversity. Even in the absence of somatic mechanisms, it is possible to generate a large number of different antibodies.

4. Junctional diversity: The DNA rearrangement mechanism that generates the V gene appears to be imprecise in that nucleotides at the ends of the gene segments can be deleted in the region of the joining event (Sakano *et al.*, 1980; Weigert *et al.*, 1980; Kurosawa *et al.*, 1981). This can lead to codon changes at the V-D-J and V-J junctions.

Because all immunoglobulin gene segments seem to be translated in only one reading frame in order to give rise to a functional immunoglobulin chain, one implication of this imprecision is that these gene segments often join in an out-of-phase reading frame. This type of nonproductive joining occurs frequently in lymphocytes (Altenburger *et al.*, 1980; Max *et al.*, 1980; Weigert *et al.*, 1980; Walfield *et al.*, 1981). Thus, diversity is achieved at the expense of some waste.

5. N-region diversity: For the heavy chain genes, extra nucleotides not encoded by the germline gene segments may be added between the rearranging gene segments as a result of the joining process (Sakano *et al.*, 1980; Alt and

Baltimore, 1982). This event is referred to as N-region diversity. Both junctional and N-region diversity alter the coding region of the variable region only at the joining point of the gene segments; this region correlates to the third hypervariable region observed in the comparisons of the protein data (see above; Wu and Kabat, 1970).

6. Somatic hypermutation: There is a mechanism that induces point mutations throughout the variable region gene and its untranslated regions at a late stage in B-cell development (Weigert *et al.*, 1970; Gershenfeld *et al.*, 1981; Kim *et al.*, 1981; Selsing and Storb, 1981; Clarke *et al.*, 1982). Those B cells with receptors that have higher affinities for antigen as a result of somatic hypermutation are believed to be selectively expanded under conditions of limiting antigen. This mechanism appears to induce point mutations randomly, and is correlated to the class of the antibody (Crews *et al.*, 1981). Therefore, μ chains tend to be encoded by germline sequences, while the other chains are often encoded by variants (Gearhart *et al.*, 1981). These observations led to the hypothesis that the somatic hypermutation event may be induced during the class switch rearrangement (Crews *et al.*, 1981; Kim *et al.*, 1981).

Immunoglobulin gene evolution

One of the most interesting questions to emerge from the analysis of the immunoglobulin gene family is the problem of immunoglobulin gene evolution. It is presumably advantageous for the organism to maintain a large repertoire of germline gene segments, yet, it is important to maintain the basic framework structure of the variable region. Therefore, there must be some selection pressure to maintain portions of the variable region (Gally and Edelman, 1972; Smith *et al.*, 1971). Analyses of the coding region of V genes have indicated that the CDR regions are mutating at a much higher rate than the framework regions

(Ohta, 1981); the nature of this high mutation rate is unclear, however, as the CDR regions are quite small, and therefore the calculation is subject to considerable error.

The problem of the T-cell antigen receptor

Like the B lymphocytes, the T lymphocytes are capable of recognizing a wide variety of antigens. Unlike B cells, however, T cells must recognize antigen in the context of membrane-bound gene products of the Major Histocompatibility Complex (MHC) (Kindred and Shreffler, 1972; Katz *et al.*, 1973; Zinkernagel and Doherty, 1974). The murine MHC encodes at least two classes of cell-surface molecules that play direct roles in T-cell antigen recognition. The class I molecules include the K, D, R and L genes which are commonly referred to as the transplantation antigens and are found on virtually all cells in an organism. The class II molecules are encoded by the I region, and include the A_β, A_α, E_β and E_α genes. These molecules are found only on B lymphocytes and the antigen-presenting cells, the macrophage. Studies of T-cell responses have shown that the T_C cells tend to recognize the class I gene products (Alter *et al.*, 1973; Zinkernagel and Doherty, 1975; 1979), while the T_H cells tend to recognize the class II molecules (Alter *et al.*, 1973; Katz *et al.*, 1975; Meo *et al.*, 1975; Thomas *et al.*, 1977; Swierkosz *et al.*, 1979). Analyses of different antigen systems have shown that the T lymphocyte, like the B lymphocyte, is capable of discerning between closely-related antigens (Barcinski and Rosenthal, 1977; Solinger *et al.*, 1979; Thomas *et al.*, 1980). As the problems inherent in antigen-recognition appear to be similar between the B and T lymphocytes, it was perhaps not unreasonable to hypothesize that immunoglobulins served as the antigen-binding receptor on T as well as B lymphocytes (Crone *et al.*, 1972; Nossal *et al.*, 1972; Avrameas *et al.*, 1979). It was reported early on that low levels of immuno-

globulin were in fact detectable on populations of T lymphocytes by immunofluorescence techniques (Raff, 1971; Nossal *et al.*, 1972; Jensenius and Williams, 1974). Other groups reported the presence of B-lymphocyte idiotypes on T cells that mapped to chromosome 12 (Binz and Wigzell, 1975, 1976; Cosenza *et al.*, 1977; Krammer and Eichmann, 1977). However, subsequent analyses using cloned immunoglobulin DNA probes clearly indicated that no immunoglobulin gene segments appear to be utilized in the generation of the T-cell antigen receptor (Kronenberg *et al.*, 1980; Kraig *et al.*, 1983; Kronenberg *et al.*, 1983; Nakanishi *et al.*, 1982; Kemp *et al.*, 1982). Other reports involving T-cell-specific alloantigens (Owen *et al.*, 1981) and antigen-binding factors secreted by T cells (Murphy, 1978; Wieder *et al.*, 1982; Kapp and Araneo, 1982; Tanaguchi *et al.*, 1982; Krupen *et al.*, 1982) appeared to represent promising leads for the T-cell antigen receptor; despite many attempts, however, further characterization of these alloantigens and factors proved to be difficult.

Clone-specific monoclonal antibodies

Starting in 1983, monoclonal antibodies were generated that bound to and inhibited only one of a panel of murine T-cell lymphomas (Allison *et al.*, 1982; McIntyre and Allison, 1983) cloned murine T_H cells (Haskins *et al.*, 1983) and cytotoxic cells (Lancki *et al.*, 1983) and human cytotoxic cells (Meuer *et al.*, 1983). In the case of the murine T_H cells, the monoclonal antibody KJ1-26 was shown to bind only a T-cell hybridoma specific for chick albumin (cOVA) and restricted to I-A^d. The monoclonal antibody was specific only for this hybridoma; no binding of the antibody could be demonstrated with any other T-cell hybrid tested or with variants of the hybridoma that lost their ability to respond to the proper antigen-MHC combination. This antibody also specifically inhibited the proliferation of the target T cell by antigen (Haskins *et al.*, 1983; Kappler *et al.*,

1983). These data indicated that these antibodies were directed against the antigen-specific region of the T-cell antigen receptor. Experiments with human cytotoxic T cell lines produced monoclonal antibodies that reacted with a large number of T cell-specific alloantigens. As in the murine T_H system, monoclonal antibodies were generated against specific T_C cell lines (Meuer *et al.*, 1983). In every case the monoclonal antibodies appeared to recognize a heterodimer consisting of an acidic α chain and a slightly basic β chain of 40,000-43,000 daltons each in mice, and an α chain of 45,000 daltons, and a β chain of 40,000 daltons in humans. Comparisons of the immunoprecipitated heterodimers using peptide maps indicated that each chain of the heterodimer consisted of a region that was variable and a region that was constant between different T cells, indicating that, like the immunoglobulin, the T-cell antigen receptor consisted of a variable region for antigen-binding, and a constant region responsible for effector functions (McIntyre *et al.*, 1983; Kappler *et al.*, 1983; Meuer *et al.*, 1984).

The isolation of T cell-specific cDNA clones

Characterization of the T-cell antigen receptor on the DNA level still proved to be elusive. Later, using different techniques, two groups isolated T cell-specific cDNA clones that shared structural homology with immunoglobulins (Hedrick *et al.*, 1984b; Yanagi *et al.*, 1984). These clones, one from human, the other from mouse, were found to be quite homologous in one portion, indicating that they probably encoded a homologous protein in the two systems. In one case, the cDNA clones were isolated by screening a T-cell cDNA library first with ^{32}P-labelled T-cell cDNA, then subsequently with ^{32}P-labelled B-cell cDNA; T-cell specific clones were identified as having hybridized with the T-cell cDNA probe, but not the B-cell cDNA probe (Yanagi *et al.*, 1984). In the other case, a T-cell cDNA library was screened with a subtracted probe. This probe consisted of ^{32}P-

labelled cDNA made from the membrane-bound polysomal RNA of T cells in which sequences expressed by B cells had been removed by RNA hybridization (Hedrick *et al.*, 1984a). The cDNA clones isolated using these two techniques were found to be expressed as 1.3 kb and 1.0 kb messages and to detect rearrangement specifically in T cells. Sequence analysis of these clones indicated that they consisted of regions that were homologous to the immunoglobulin V, J and C regions (Hedrick *et al.*, 1984a; 1984b; Yanagi *et al.*, 1984). These data indicated that these cDNAs may in fact encode one of the chains of the T-cell antigen receptor. Amino acid sequence analysis of the β chain of the human T-cell receptor was found to match the translated sequence of one of the cDNA clones, indicating that this was indeed the case (Acuto *et al.*, 1984).

We were interested in the structure, the generation of diversity, and the evolution of the antigen receptors of lymphocytes. The first chapter of this thesis describes a study on the structure and evolution of a small subfamily of germline V_H gene segments, called the T15 family (Crews *et al.*, 1981). Specifically, we were interested in studying the flanking regions of genomic V gene segments for possible transcriptional or rearrangement control sequences, and the coding regions for indications on how these gene segments are evolving. The remainder of this thesis describes studies to characterize the variable region of the β chain of the T-cell antigen receptor in mice and humans. We were interested in studying the structure and evolution of the genes that encode the T-cell antigen receptor and the mechanisms in which diversity is generated.

REFERENCES

Acuto, O., Fabbi, M., Smart, J., Poole, C. B., Protentis, J., Royer, H. D., Schlossman, S. F., and Reinherz, E. L. (1984). Purification and N-terminal amino acid sequencing of a human T-cell antigen receptor. Proc. Natl. Acad. Sci. USA **81**, 3851-3855.

Allison, J. P., McIntyre, B. W., and Bloch, D. (1982). Tumor-specific antigen of murine T-lymphoma defined with monoclonal antibody. J. Immunol. **129**, 2293-2300.

Alt, F., Bothwell, A., Knapp, M., Siden, E., Mather, E., Koshland, M., and Baltimore, D. (1980). Synthesis of secreted and membrane-bound immunoglobulin Mu heavy chains is directed by mRNAs that differ at their 3' ends. Cell **20**, 293-301.

Alt, F., and Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: Implications from a chromosome with evidence of three D-J_H fusions. Proc. Natl. Acad. Sci. U.S.A. **78**, 5812-5816.

Altenburger, W., Steinmetz, M., and Zachau, H. G. (1980). Functional and non-functional joining in immunoglobulin light chain genes of a mouse myeloma. Nature **287**, 603-607.

Altenburger, W., Neumaier, P., Steinmetz, M., and Zachau, H. (1981). DNA sequence of the mouse immunoglobulin Kappa chain. Nucl. Acids Res. **9**, 971-981.

Alter, B. J., Schendel, D. J., Bach, M. L., Bach, F. H., Klein, J. and Stimpfling, J. H. (1973). Cell-mediated lympholysis. Importance of serologically defined H-2 regions. J. Exp. Med. **137**, 1303-1309.

Amzel, L., Poljack, R., Saul, F., Varga, J., and Richards, F. (1974). The three-dimensional structure of a combining-site ligand complex of immunoglobulin NEW at 3.5A resolution. Proc. Natl. Acad. Sci. U.S.A. **71**, 1427-1430.

Appella, E. (1971). Amino acid sequences of two mouse immunoglobulin Lambda chains. Proc. Natl. Acad. Sci. U.S.A. **68**, 590-594.

Avrameas, S., Hösli, P. Stanislawski, M. Rodrigot, M. and Vogt, E. (1979). A quantitative study at the single cell level of immunoglobulin antigenic determinants present on the surface of murine B and T lymphocytes. J. Immunol. **122**, 648-659.

Barcinski, M. A., and Rosenthal, A. S. (1977). Immune response gene control of determinant selection. I. Intramolecular mapping of the immunogenic sites on insulin recognized by guinea pig T and B cells. J. Exp. Med. **145**, 726-742.

Battey, J., Max, E., McBride, W., Swan, D., and Leder, P. (1982). A processed human immunoglobulin ξ gene has moved to chromosome 9. Proc. Natl. Acad. Sci. U.S.A. **79**, 5956-5960.

Bentley, D., and Rabbitts, T. (1980). Human immunoglobulin variable region genes - DNA sequences of two V_{κ} genes and pseudogene. Nature **288**, 730-733.

Bernard, O., Hozumi, N., and Tonegawa, S. (1978). Sequences of mouse immunoglobulin light chain genes before and after somatic changes. Cell **15**, 1133-1144.

Bernard, O., and Gough, N. (1980). Nucleotide sequence of immunoglobulin heavy chain joining segments between translocated V_H and μ constant region genes. Proc. Natl. Acad. Sci. U.S.A. **77**, 3630-3634.

Binz, H., and Wigzell, H. (1975). Shared idiotypic determinants on B and T lymphocytes reactive against the same antigenic determinants. I. Demonstration of similar or identical idiotypes on IgG molecules and T-cell receptors with specificity for the same alloantigens. J. Exp. Med. **142**, 197-211.

Binz, H., Wigzell, H., and Bazin, H. (1976). T-cell idiotypes are linked to immunoglobulin heavy chain genes. Nature **264**, 639-642.

Blomberg, B., Traunecker, A., Eisen, H., and Tonegawa, S. (1981). Organization of four mouse λ light chain immunoglobulin genes. Proc. Natl. Acad. Sci. U.S.A. **78**, 3765-3769.

Blomberg, B., and Tonegawa, S. (1982). DNA sequences of the joining regions of mouse λ light chain immunoglobulin genes. Proc. Natl. Acad. Sci. U.S.A. **79**, 530-533.

Brack, C., Hirama, M., Lenhard-Schuller, R., and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. Cell **15**, 1-14.

Brodeur, P., and Riblet, R. (1984). The immunoglobulin heavy chain variable region (Igh-V) locus in mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. Eur. J. Immunol. **14**, 922-930.

Cantor, H., and Boyse E. A. (1975). Functional subclasses of T lymphocytes bearing different Ly antigens. II. Cooperation between subclasses of Ly^+ cells in the generation of killer activity. J. Exp. Med. **141**, 1390-1399.

Capra, J., and Kehoe, J. (1974). Variable region sequences of five human immunoglobulin heavy chains of the V_H III subgroup: Definitive identification of four heavy chain hypervariable regions. Proc. Natl. Acad. Sci. U.S.A. **71**, 845-849.

Cerottini, J. C., Nordin, A. A., and Brunner, K. T. (1970). Specific *in vitro* cytotoxicity of thymus-derived lymphocytes sensitized to alloantigens. Nature **228**, 1308-1309.

Cheng, H.-L., Blattner, F., Fitzmaurice, L., Mushinski, J., and Tucker, P. (1982). Structure of genes for membrane and secreted murine IgD heavy chains. Nature **296**, 410-415.

Clarke, C., Berenson, J. Goverman, J., Boyer, P., Crews, S., Siu, G., and Calame, K. (1982). An immunoglobulin promoter region is unaltered by DNA rearrangement and somatic mutation during B-cell development. Nuc. Acids Res. **10**, 7731-7749.

Cohn, M., Blomberg, B., Geckeler, W., Raschke, W., Riblet, R., and Weigert, M. (1974). First order considerations in analyzing the generation of diversity in the immune system: Genes, receptors, signals (ed. E. E. Sercarz *et al.*), p. 89, Academic Press, New York.

Cory, S., Tyler, B., and Adams, J. (1981). Sets of immunoglobulin V_κ genes homologous to ten cloned V_κ sequences: Implications for the number of germline V_κ genes. J. Mol. Applied Genet. **1**, 103-116.

Cosenza, H., Julius, M. H., Augustin, A. A. (1977). Idiotypes as variable region markers: Analogies between receptors on phosphorylcholine-specific T and B lymphocytes. Immunol. Rev. **34**, 3-33.

Crews, S., Griffin, J., Huang, H., Calame, K., and Hood, L. (1981). A single V_H gene segment encodes the immune response to phosphorylcholine: Somatic mutation is correlated with the class of the antibody. Cell **25**, 59-66.

Crone, M., Koch, C., and Simonsen, M. (1972). The elusive T cell receptor. Transpl. Rev. **10**, 36-56.

Davis, M., Calame, K., Early, P., Livant, D., Joho, R., Weissman, I., and Hood, L. (1980a). An immunoglobulin heavy-chain gene is formed by at least two recombinational events. Nature **283**, 733-739.

Davis, M., Kim, S., and Hood, L. (1980b). Immunoglobulin class switching: Developmentally regulated DNA rearrangements during differentiation. Science **209**, 1360-1365.

Dildrop, R. (1984). A new classification of mouse V_H sequences. Immunol. Today **5**, 85-86.

Dugan, E., Bradshaw, R., Simms, E., and Eisen, H. (1973). Amino acid sequence of the light chain of a mouse myeloma protein (MOPC-315). Biochemistry **12**, 5400-5416.

Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980a). An immunoglobulin heavy chain variable region is generated from three segments of DNA: V_H, D and J_H. Cell **19**, 981-992.

Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R., and Hood, L. (1980b). Two mRNAs can be produced from a single immunoglobulin μ gene by alternative RNA processing pathways. Cell **20**, 313-319.

Elliot, B., Steiner, L., and Eisen, H. (1981). Amino acid sequence variation in mouse $\lambda 2$ chains. Fed. Proc. **40**, 1098.

Ellison, J., Buxbaum, J., and Hood, L. (1981). Nucleotide sequence of a human immunoglobulin $C_{\gamma 4}$ gene. DNA **1**, 11-18.

Ellison, J., and Hood, L. (1982). Linkage and sequence homology of two human immunoglobulin γ heavy chain constant region genes. Proc. Natl. Acad. Sci. U.S.A. **79**, 1984-1988.

Ellison, J., Berson, B., and Hood, L. (1982). The nucleotide sequence of a human $C_{\gamma 1}$ gene. Nucl. Acids Res. **10**, 4071-4079.

Flanagan, J., and Rabbitis, T. (1982). The sequence of a human immunoglobulin epsilon heavy chain constant region gene, and evidence for three non-allelic genes. EMBO J. **1**, 655-660.

Gally, J. A., and Edelman, G. M. (1972). The genetic control of immunoglobulin synthesis. Ann. Rev. Genetics **6**, 1-45.

Gearhart, P., Johnson, N., Douglas, R., and Hood, L. (1981). IgG antibodies to phosphorylcholine exhibit more diversity than their IgM counterparts. Nature **291**, 29-34.

Gershenfeld, H. K., Tsukamoto, A., Weissman, I. L., and Joho, R. (1981). Somatic diversification is required to generate the V_{κ} gene of MOPC 511 and MOPC 167 myeloma proteins. Proc. Natl. Acad. Sci. U.S.A. **78**, 7674-7678.

Gershon, R. K. (1974). T cell suppression. Contemp. Topics in Immunobiol. **3**, 1-40.

Gough, N., and Bernard, O. (1981). Sequences of the joining region genes for immunoglobulin heavy chains and their role in generation of antibody diversity. Proc. Natl. Acad. Sci. U.S.A. **78**, 509-513.

Grant, J. A., and Hood, L. (1971). N-terminal analysis of normal immunoglobulin light chains. I. A study of thirteen individual humans. Immunochemistry **8**, 63-79.

Haskins, K., Kubo, R., White, J., Pigeon, M., Kappler, J., and Marrack, P. (1983). The major histocompatibility complex-restricted antigen receptor on T cells. I. Isolation with a monoclonal antibody. J. Exp. Med. **157**, 1149-1169.

Hedrick, S. M., Cohen, D. I., Nielsen, E. A., and Davis, M. M. (1984a). Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature **308**, 149-153.

Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I., and Davis, M. M. (1984b). Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. Nature **308**, 153-158.

Hieter, P., Maizel, J., and Leder, P. (1982). Evolution of human immunoglobulin κ J region genes. J. Biol. Chem. **257**, 1516-1522.

Hilschmann, N., and Craig, L. (1965). Amino acid sequence studies with Bence-Jones proteins. Proc. Natl. Acad. Sci. U.S.A. **53**, 1403-1409.

Honjo, T., Obata, M., Yamawaki-Katoaka, Y., Katoaka, T., Takahashi, N., and Mano, Y. (1979). Cloning and complete nucleotide sequence of mouse immunoglobulin γ1 chain gene. Cell **18**, 559-568.

Hood, L., Gray, W., Sanders, E., and Dreyer, W. (1968). Light chain evolution. Cold Spring Harbor Symp. Quant. Biol. **32**, 133-146.

Hood, L., Loh, E., Hubert, J., Barstad, P., Eaton, B., Early, P., Fuhrman, J., Johnson, N., Kronenberg, M., and Schilling, J. (1976). The structure and genetics of mouse immunoglobulins: An analysis of NZB myeloma proteins and sets of BALB/c myeloma proteins binding particular haptens. Cold Spring Harbor Symp. Quant. Biol. **41**, 817-836.

Hozumi, N., and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc. Natl. Acad. Sci. U.S.A. **73**, 3628-3632.

Ishida, N., Ueda, S., Hayashida, H., Miyata, T., and Honjo, T. (1982). The nucleotide sequence of the mouse immunoglobulin epsilon gene: comparison with the human epsilon gene sequence. EMBO J. **1**, 1117-1123.

Jensenius, J. C., and Williams, A. (1974). The binding of anti-immunoglobulin antibodies to rat thymocytes and thoracic duct lymphocytes. Eur. J. Immunol. **4**, 91-97.

Kapp, J. A., and Araneo, B.A. (1982). Characterization of antigen-specific suppressor factors from T cell hybridomas. In: Isolation, Characterization and Utilization of T Lymphocyte Clones (ed. C. G. Fathman and F. W. Fitch), pp. 137-148, Academic Press, New York.

Katoaka, T., Kawakami, T., Takahashi, N., and Honjo, T. (1980). Rearrangement of immunoglobulin γ1-chain gene and mechanism for heavy-chain class switch. Proc. Natl. Acad. Sci. U.S.A. **77**, 919-923.

Katz, D. H., Hamaoka, T., and Benacerraf, B. (1973). Cell interactions between histoincompatible T and B lymphocytes. II. Failure of physiologic cooperative interactions between T and B lymphocytes form allogeneic donor strains in humoral response to hapten-protein conjugates. J. Exp. Med. **137**, 1405-1418.

Katz, D. H., Graves, M., Dorf, M. E., Dimuzio, H., and Bernacerraf, B. (1975). Cell interactions between histoincompatible T and B lymphocytes. VII. Cooperative responses between lymphocytes are controlled by genes in the I region of the H-2 complex. J. Exp. Med. **141**, 263-268.

Kemp, D. J., Adams, J. M., Mottram, P. L., Thomas, W. R., Walker, I. D., and Miller, J. F. A. P. (1982). A search for messenger RNA molecules bearing immunoglobulin V_H nucleotide sequences in T cells. J. Exp. Med. **156**, 1848-1853.

Kim, S., Davis, M., Sinn, E., Patten, P., and Hood, L. (1981). Antibody diversity: somatic hypermutation of rearranged V_H genes. Cell **27**, 573-581.

Kindred, B., and Shreffler, D. C. (1972). H-2 dependence of co-operation between T and B cells *in vivo*. J. Immunol. **109**, 940-943.

Kraig, E., Kronenberg, M., Kapp, J. A., Pierce, C. W., Abruzzini, A. F., Sorensen, C. M., Samelson, L. E., Schwartz, R. H., and Hood, L. E (1983). T and B cells that recognize the same antigen do not transcribe similar heavy chain variable region gene segments. J. Exp. Med. **158**, 192-209.

Krammer, P. and Eichmann, K. (1977). T cell receptor idiotypes are controlled by genes in the heavy chain linkage group and the major histocompatibility complex. Nature **270**, 733-735.

Kronenberg, M., Davis, M. M., Early, P. W., Hood, L. E., and Watson, J. D. (1980). Helper and killer T cells do not express B cell immunoglobulin joining and constant region gene segments. J. Exp. Med. **152**, 1745-1761.

Kronenberg, M., Kraig, E., Siu, G., Kapp, J. A., Kappler, J., Marrack, P., Pierce, C. W., and Hood, L. (1983). Three T-cell hybridomas do not express detectable heavy chain variable gene transcripts. J. Exp. Med. **158**, 210-227.

Krupen, K., Araneo, B. A., Brink, L., Kapp, J. A., Stein, S., Wieder, K. J., and Webb, D. R. (1982). Purification and characterization of a monoclonal T cell suppressor factor specific for L-glutamic acid^{60}-L-alanine^{30}-L-tyrosine^{10} (GAT). Proc. Natl. Acad. Sci. U.S.A. **79**, 1254-1258.

Kurosawa, Y., von Boehmer, H., Haas, W., Sakano, H., Trauneker, A., and Tonegawa, S. (1981). Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. Nature **290**, 565-570.

Lancki, D. W., Lorber, M. I., Loken, M. R., and Fitch, F. W. (1983). A clone-specific monoclonal antibody that inhibits cytolysis of a cytolytic T cell clone. J. Exp. Med. **157**, 921-935.

Lennox, E., and Cohn, M. (1967). Immunoglobulins. Ann. Rev. Biochem. **36**, 365-406.

Maki, R., Traunecker, A., Sakano, H., Roeder, W., and Tonegawa, S. (1980). Exon shuffling generates an immunoglobulin heavy chain gene. Proc. Natl. Acad. Sci. U.S.A. **77**, 2138-2142.

Max, E., Seidman, J., and Leder, P. (1979). Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region gene. Proc. Natl. Acad. Sci. U.S.A. **76**, 3450-3454.

Max, E. E., Seidman, J. G., Miller, H. and Leder, P. (1980). Variation in the crossover point of kappa immunoglobulin gene V-J recombination: Evidence from a cryptic gene. Cell **21**, 793-799.

Max, E., Maizel, J., and Leder, P. (1981). The nucleotide sequence of a 5.5 kilobase DNA segment containing the mouse κ immunoglobulin J and C region genes. J. Biol. Chem. **256**, 5116-5120.

Max, E., Battey, J., Ney, R., Kirsch, I., and Leder, P. (1982). Duplication and deletion in the human immunoglobulin ξ genes. Cell **29**, 691-699.

McIntyre, B. W., and Allison, J. P. (1983). The mouse T cell receptor: Structural heterogeneity of molecules of normal T cells defined by xenoantiserum. Cell **34**, 739-746.

Meo, T., David, C. S., Rijnbeek, H. M., Nabholz, M., Miggiano, V. C., and Shreffler, D. C. (1975). Inhibition of mouse MLR by anti-Ia sera. Transpl. Proc. **7**, 127-129.

Meuer, S. C., Fitzgerald, K. A., Hussey, R. E., Hodgdon, J. C., Schlossman, S. F., and Reinherz, E. L. (1983). Clonotypic structure involved in antigen-specific human T cell function. Relationship to the T3 molecular complex. J. Exp. Med. **157**, 705-719.

Meuer, S. C., Acuto, O., Hercend, T., Schlossman, S. F., and Reinherz, E. L. (1984). The human T-cell receptor. Ann. Rev. Immunol. **2**, 23-50.

Miller, J., Selsing, E., and Storb, U. (1982). Structural alterations in J regions of mouse immunoglobulin λ genes are associated with differential gene expressions. Nature **295**, 428-430.

Milstein, C. (1967). Linked groups of residues in immunoglobulin Kappa chains. Nature **216**, 330-332.

Mitchell, G. F., and Miller, J. F. A. P. (1968). Cell-to-cell interaction in the immune response. II. The source of hemolysin-forming cells in irradiated mice given bone marrow and thymus or thoracic duct lymphocytes. J. Exp. Med. **128**, 821-837.

Murphy, D. B. (1978). The I-J subregion of the murine H-2 gene complex. Springer Seminars Immunopathol. **1**, 111-130.

Nakanishi, K., Sugimura, K., Yaoita, Y., Maeda, K., Kashiwamura, S-I., Honjo, T., and Kishimoto, T. (1982). A T15-idiotype-positive T suppressor hybridoma does not use the T15 V_H gene segment. Proc. Natl. Acad. Sci. U.S.A. **79**, 6984-6988.

Nishida, Y., Miki, T., Hisajima, H., and Honjo, T. (1982). Cloning of human immunoglobulin ξ chain genes. Evidence for multiple C_ξ genes. Proc. Natl. Acad. Sci. U.S.A. **79**, 3833-3837.

Nossal, G. J. V., Warner, N. L., Lewis, H., and Sprent, J. (1972). Quantitative features of a sandwich radioimmunolabelling technique for lymphocyte surface receptors. J. Exp. Med. **135**, 405-428.

Ohta, T. (1981). In: Evolution and variation of multigene families, (ed. S. Levin), p. 72-87, Springer Verlag, Berlin.

Ollo, R., Auffray, C., Morchamps, C., and Rougeon, F. (1981). Comparison of mouse immunoglobulin γa and γb chain genes suggests that exons can be exchanged between genes in a multigenic family. Proc. Natl. Acad. Sci. U.S.A. **78**, 2442-2446.

Owen, F. L., Riblet, R., and Taylor, B. A. (1981). The T suppressor cell alloantigen Tsu^d maps near immunoglobulin allotype genes and may be a heavy chain constant region marker on a T cell receptor. J. Exp. Med. **153**, 801-810.

Padlan, E., Segal, D., Spande, T., Davies, D., Rudikoff, S., and Potter, M. (1973). Structure at 4.5 Å resolution of a phosphorycholine-binding Fab. Nature New Biol. **245**, 165-167.

Putnam, M., Newell, J., Rudikoff, S., and Haber, E. (1971). Classification of mouse V_κ groups based on the partial amino acid sequence to the first invariant tryptophan: Impact of 14 new sequences from IgG myeloma proteins. Mol. Immunol. **19**, 1619-1630.

Rabbitts, T., Forster, A., and Milstein, C. (1981). Human immunoglobulin heavy chain genes: Evolutionary comparisons of Cμ, Cδ and Cγ genes and associated switch sequences. Nucl. Acids Res. **9**, 4509-4524.

Raff, M. C. (1971). Surface antigenic markers for distinguishing T and B lymphocytes in mice. Transpl. Rev. **6**, 52-80.

Ravetch, J., Siebenlist, U., Korsmeyer, S., Waldmann, T., and Leder, P. (1981). Structure of the human immunoglobulin μ locus: Characterization of embryonic and rearranged J and D genes. Cell **27**, 583-591.

Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L., and Wall, R. (1980). Two mRNAs with different 3' ends encode membrane-bound and secreted forms of immunoglobulin μ chain. Cell **20**, 303-312.

Rothenberg, E. and Lugo, J. Differentiation and cell division in the mammalian thymus. Devel. Biol., in press.

Sakano, H., Huppi, K., Heinrich, G., and Tonegawa, S. (1979). Sequences of the somatic recombination sites of immunoglobulin light-chain genes. Nature **280**, 288-294.

Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., and Tonegawa, S. (1980). Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. Nature **286**, 676-683.

Sakano, H., Kurosawa, Y., Weigert, M., and Tonegawa, S. (1981). Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes. Nature **290**, 562-565.

Schilling, J., Clevinger, B., Davie, J. M., Hood, L. (1980) Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V-region gene segments. Nature **283**, 35-40.

Scollay, R., Bartlett, P., and Shortman, K. (1984). T cell development in the adult murine thymus: Changes in the expression of the surface antigens Ly2, L3T4 and B2A2 during development from early precursor-cells to emigrants. Immunol. Rev. **82**, 79-103.

Segal, D., Padlan, E., Cohen, G., Rudikoff, S., Potter, M., and Davies, D. (1974). The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. Proc. Natl. Acad. Sci. U.S.A. **71**, 4298-4302.

Seidman, J., Max, E., and Leder, P. (1979). A κ-immunoglobulin gene is found by site-specific recombination without further somatic mutation. Nature **280,** 370-375.

Selsing, E., and Storb, U. (1981). Somatic mutation of immunoglobulin light-chain variable-region genes. Cell **25,** 47-58.

Shimizu, A., Takahashi, N., Yamawaki-Katoaka, Y., Nishida, Y., Katoaka, T., and Honjo, T. (1981). Ordering of mouse immunoglobulin heavy chain genes by molecular cloning. Nature **289,** 49-53.

Siebenlist, U., Ravetch, J., Korsmeyer, S., Waldmann, T., and Leder, P. (1981). Human immunoglobulin D segments encoded in tandem multigenic families. Nature **294,** 632-635.

Smith, G. P., Hood, L., and Fitch, W. M. (1971). Antibody diversity. Ann. Rev. Biochem. **40,** 969-1012.

Smithies, O. (1967). The genetic basis of antibody variability. Cold Spring Harb. Symp. Quant. Biol. **32,** 161-168.

Solinger, A. M., Ultec, M. E., Margoliash, E., and Schwartz, R. H. (1979). T-lymphocyte response to cytochrome c. I. Demonstration of a T-cell heteroclitic proliferative response and identification of a topographic antigenic determinant on pigeon cytochrome c whose immune recognition requires two complementing major histocompatibility complex-linked immune response genes. J. Exp. Med. **150,** 830-848.

Swierkosz, J. E., Marrack, P., and Kappler, J. W. (1979). The role of H-2 linked genes in helper T cell function. V. I-region control of helper T cell interaction with antigen-presenting macrophages. J. Immunol. **123,** 654-659.

Takahashi, N., Ueda, S., Obata, M., Nikaido, T., Nakai, S., and Honjo, T. (1982). Structure of human immunoglobulin gamma genes: Implications for the evolution of a gene family. Cell **29,** 671-679.

Taniguchi, M., Tokuhisa, T., Kanno, M., Yaoita, Y. Shimizu, A., and Honjo, T. (1982). Reconstitution of antigen-specific suppressor activity with translation products of mRNA. Nature **298**, 172-174.

Thomas, D. W., Kamashita, U., and Shevach, E. M. (1977). Nature of the antigenic complex recognized by T lymphocytes. IV. Inhibition of antigen-specific proliferation by antibodies to stimulator macrophage Ia antigens. J. Immunol. **119**, 223-226.

Thomas, D. W., Hsieh. K-H., Schauster, J. L., Mudd, M. S., and Wilner, G. D. (1980). Nature of T lymphocyte recognition of macrophage-associated antigens. V. Contribution of individual peptide residues of human fibrino peptide B to T lymphocyte responses. J. Exp. Med. **152**, 620-632.

Tonegawa, S., Maxam, A., Tizard, R., Bernard, O., and Gilbert, W. (1978). Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc. Natl. Acad. Sci. U.S.A. **75**, 1405-1489.

Tucker, P., Marcu, K., Newell, N., Richards, J., and Blattner, F. (1979). Sequence of the cloned gene for the constant region of the murine γ2b immunoglobulin heavy chain. Science **206**, 1303-1306.

Tucker, P., Slightom, J., and Blattner, F. (1981). Mouse IgA heavy chain gene sequence: Implications for evolution of immunoglobulin hinge exons. Proc. Natl. Acad. Sci. U.S.A. **78**, 7684-7688.

Ueda, S., Nakai, S., Nishida, Y., Hisajima, H., and Honjo, T. (1982). Long terminal repeat-like elements flank a human immunoglobulin epsilon pseudogene that lacks introns. EMBO J. **1**, 1539-1544.

Walfield, A., Selsing, E., Arp, B., and Storb, U. (1981). Misalignment of V and J gene segments resulting in a nonfunctional immunoglobulin gene. Nucl. Acid Res. **9**, 1101-1109.

Weigert, M. G., Cesari, I. M., Vonkovich, S. J., and Cohn, M. (1970). Variability in the lambda light chain sequence of mouse antibody. Nature **228**, 1045-1047.

Weigert, M., and Riblet, R. (1976). Genetics of variable regions. Cold Spring Harb. Symp. Quant. Biol. **41**, 837-846.

Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J., Hood, L. (1980). The joining of V and J gene segments creates antibody diversity. Nature **283**, 497-499.

Whitlock, C., Denis, K., Robertson, D., and Witte, O. (1985). *In vitro* analysis of murine B-cell development. Ann. Rev. Immunol. **3**, 213-235.

Wieder, K. J., Araneo, B. A., Kapp, J. A., and Webb, D. R. (1982). Cell-free translation of a biologically active, antigen-specific suppressor T cell factor. Proc. Natl. Acad. Sci. U.S.A. **79**, 3599-3603.

Wu, G., Govindji, N., Hozumi, N., and Murialdo, H. (1982). Nucleotide segments of a chromosomal rearranged λ2 immunoglobulin gene of mouse. Nucl. Acids Res. **19**, 3831-3843.

Wu, T., and Kabat, E. (1970). Analysis of the sequences of Bence-Jones proteins and myeloma light chains and their implications of antibody complementarity. J. Exp. Med. **132**, 211-250.

Yamawaki-Katoaka, Y., Katoaka, T., Takahashi, N., Obata, M., and Honjo, T. (1980). Complete nucleotide sequence of immunoglobulin γ2b chain gene cloned from mouse DNA. Nature **283**, 786-789.

Yamawaki-Katoaka, Y., Miyata, T., and Honjo, T. (1981). The complete nucleotide sequence of mouse immunoglobulin γ2a gene and evolution of heavy chain genes: Further evidence for intervening sequence-mediated domain transfer. Nucl. Acids Res. **9**, 1365-1381.

Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S. P., Aleksander, I., and Mak, T. W. (1984). A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. Nature **308**, 145-149.

Zeelon, E., Bothwell, A., Kantor, F., and Schechter, I. (1981). An experimental approach to enumerate the genes coding for immunoglobulin variable-regions. Nucleic Acids Res. **9**, 3809-3820.

Zinkernagel, R. M., and Doherty, P. C. (1974). Restriction of *in vitro* T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semi allogeneic system. Nature **248**, 701-702.

Zinkernagel, R. M., and Doherty, P. C. (1975). H-2 compatibility requirements for T-cell-mediated lysis of target cells infected wih lymphocytic choriomeningitis virus. J. Exp. Med. **141**, 1427-1436.

Zinkernagel, R. M., and Doherty, P. C. (1979). MHC restricted cytotoxic T cells: Studies on the biological role of polymorphic major transplanation antigens determining T cell restriction specificity, function and responsiveness. Adv. Immunol. **27**, 51-177.

Figure 1. An immunoglobulin molecule. The amino and carboxy terminals and interchain cystine disulfide bridges are indicated. VL and CL represent the variable and constant regions of the light chain. VH represents the variable regions of the heavy chain; CH_1, CH_2, CH_3 and CH_4 represent the domains ofthe heavy chain constant region. H indicates the hinge region of the heavy chain. The variable regions are subdivided into seven sections: four framework regions denoted FR; and the complementarity-determining regions, or hypervariable regions, denoted CDR. The portions of the heavy chain variable region that are encoded by the D_H and J_H gene segments and the portion of the light chain variable region that is encoded by the J_L gene segment are indicated.

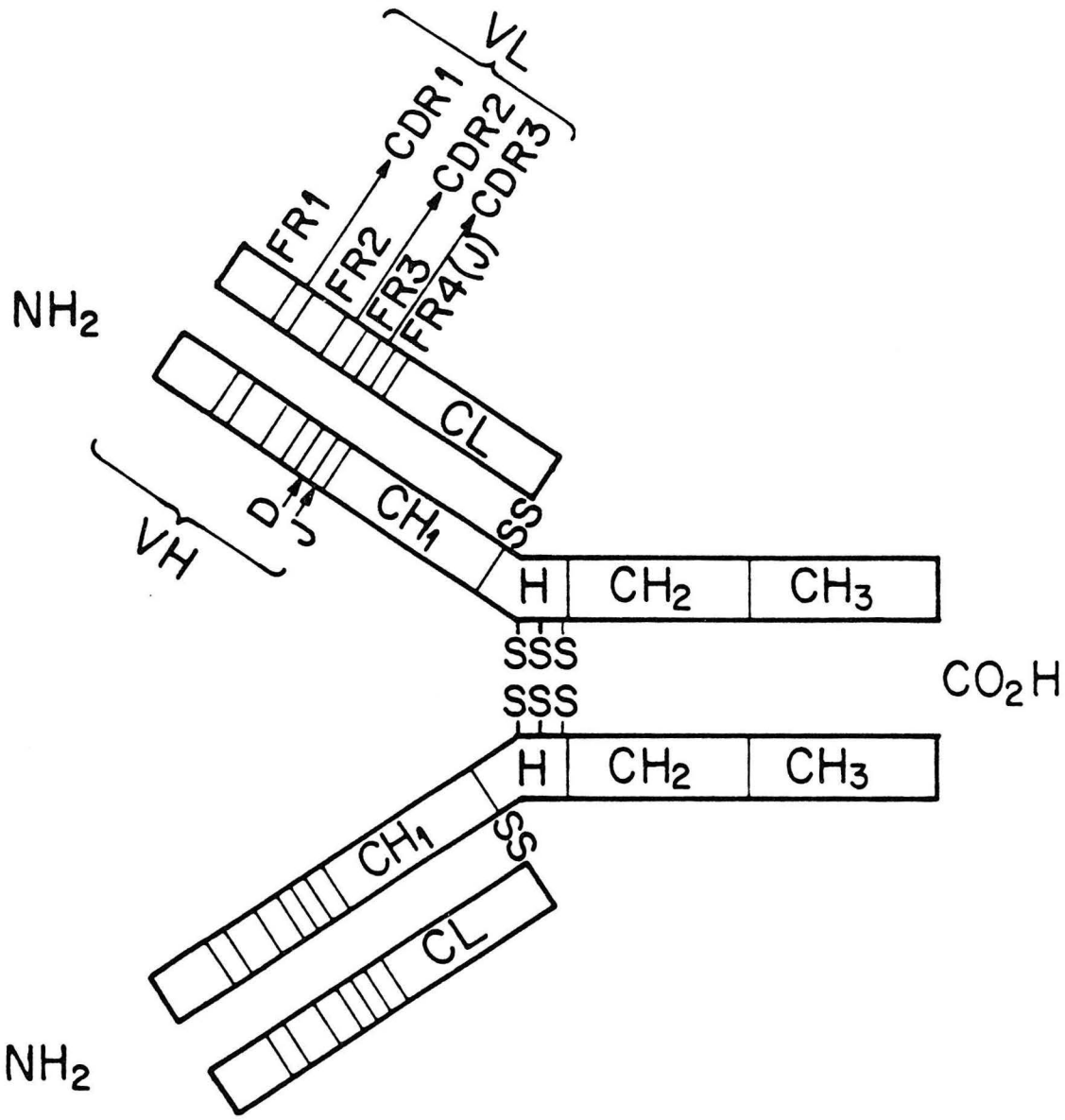

VL
CDR1
CDR2
CDR3
FR1
FR2
FR3
FR4(J)
NH2
CL
CH1
D
J
VH
SS
H
CH2
CH3
SSS
SSS
CO2H
H
CH2
CH3
CH1
SS
CL
NH2

Figure 2. Immunoglobulin gene rearrangement and mRNA splicing. A) Light chain. The V_L and J_L gene segments and the C gene are indicated; L is the hydrophobic leader of the V_L gene, and P denotes the initiation point of transcription. DNA rearrangement involves the deletion of the DNA between the V and J gene segments. Splicing of the hnRNA removes the introns between the leader and the variable region gene and the variable region gene and the constant region gene. B) Heavy chain. The V_H, D_H, and J_H gene segments and the C_μ gene are indicated. The L is the leader, and P denotes the initiation point of transcription. The exons of the C_μ gene are indicated 1-4, with S indicating the secreted exon, and M indicating the membrane-bound exon. The hatched boxes indicate 3' untranslated regions. The rearrangement and splicing events are similar to those of the light chain genes except that a differential splicing event at the 3' end of the gene generates the potential for two carboxy termini: one that serves as the membrane anchor of the membrane-bound immunoglobulin, the other as a secreted tail.

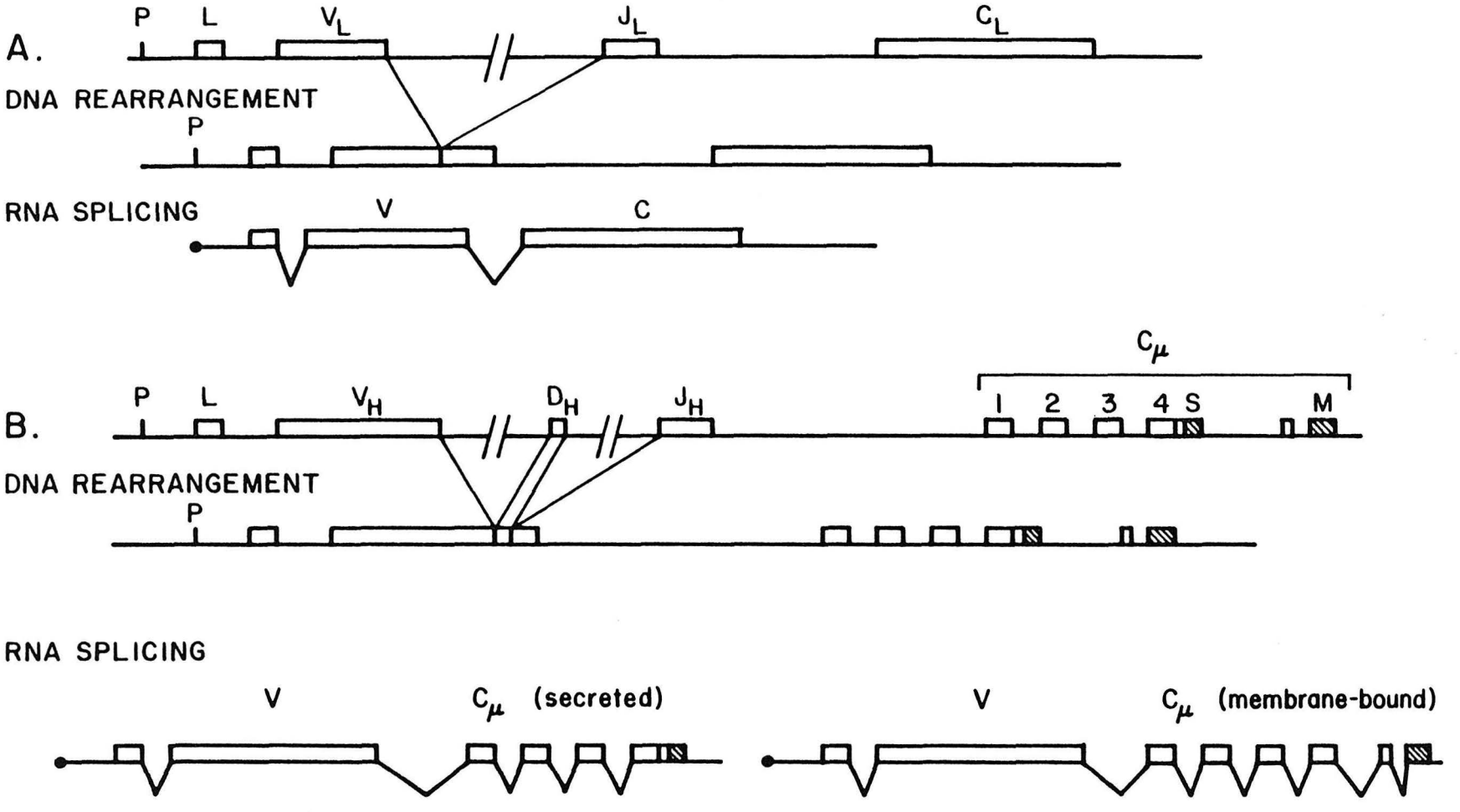

A.
P
L
V_L
J_L
C_L
DNA REARRANGEMENT
P
RNA SPLICING
V
C
B.
C_μ
P
L
V_H
D_H
J_H
1
2
3
4
S
M
DNA REARRANGEMENT
P
RNA SPLICING
V
C_μ (secreted)
V
C_μ (membrane-bound)

Figure 3. Genomic organization of the immunoglobulin gene families. Crosshatches indicate that the physical linkage has not been determined.

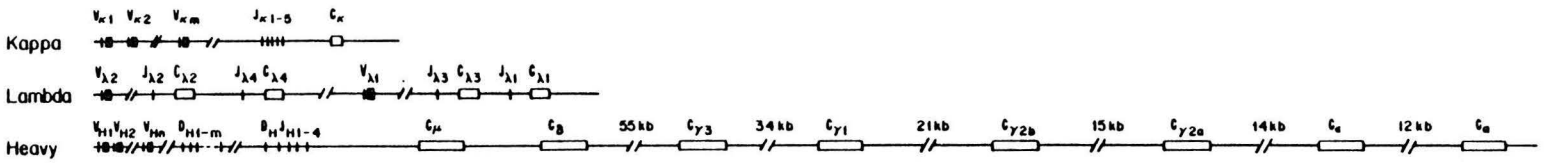

Kappa
Lambda
Heavy
55kb
34kb
21kb
15kb
14kb
12kb

Figure 4. A model for the heavy-chain class switch from C_μ to C_α. The variable region gene and the constant region genes are indicated. S_μ, S_γ and S_α indicate the switch regions for the C_μ, C_γ and C_α genes. P_μ and P_α indicate putative switch proteins. Rearrangement during the heavy chain class switch occurs within the switch regions.

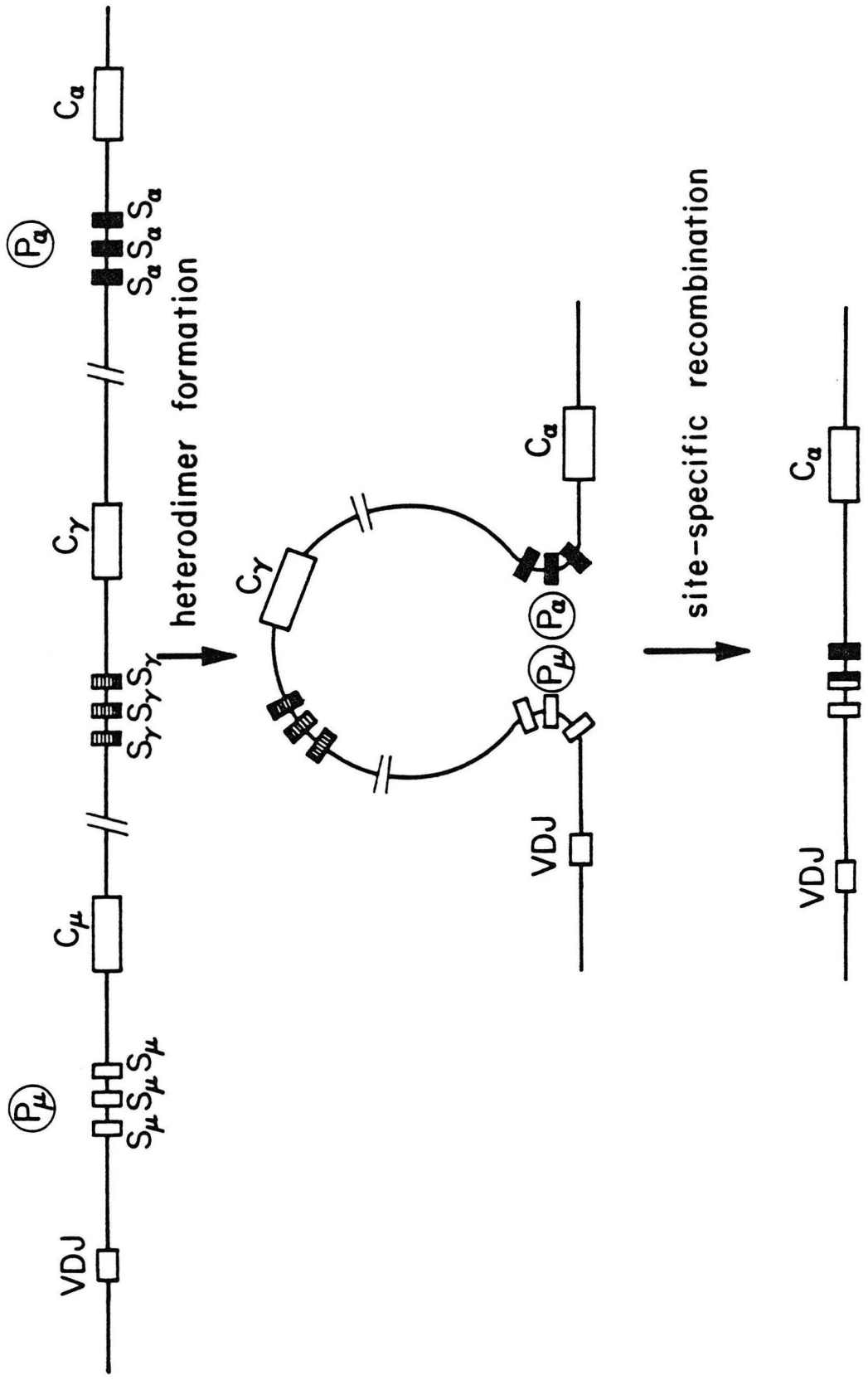
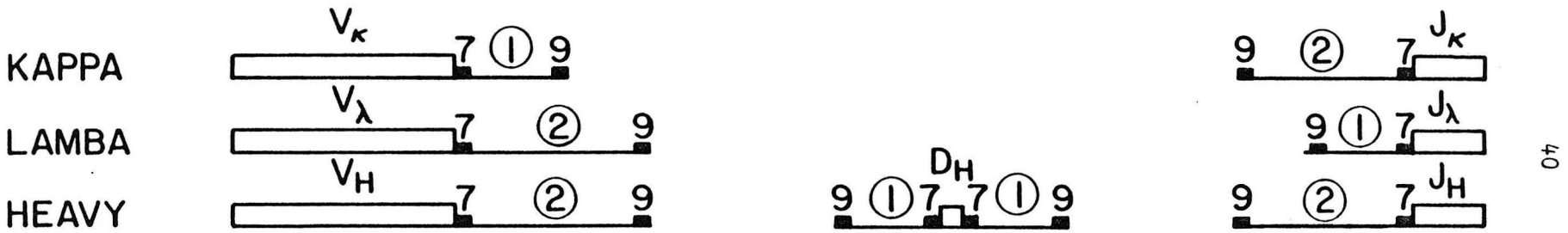
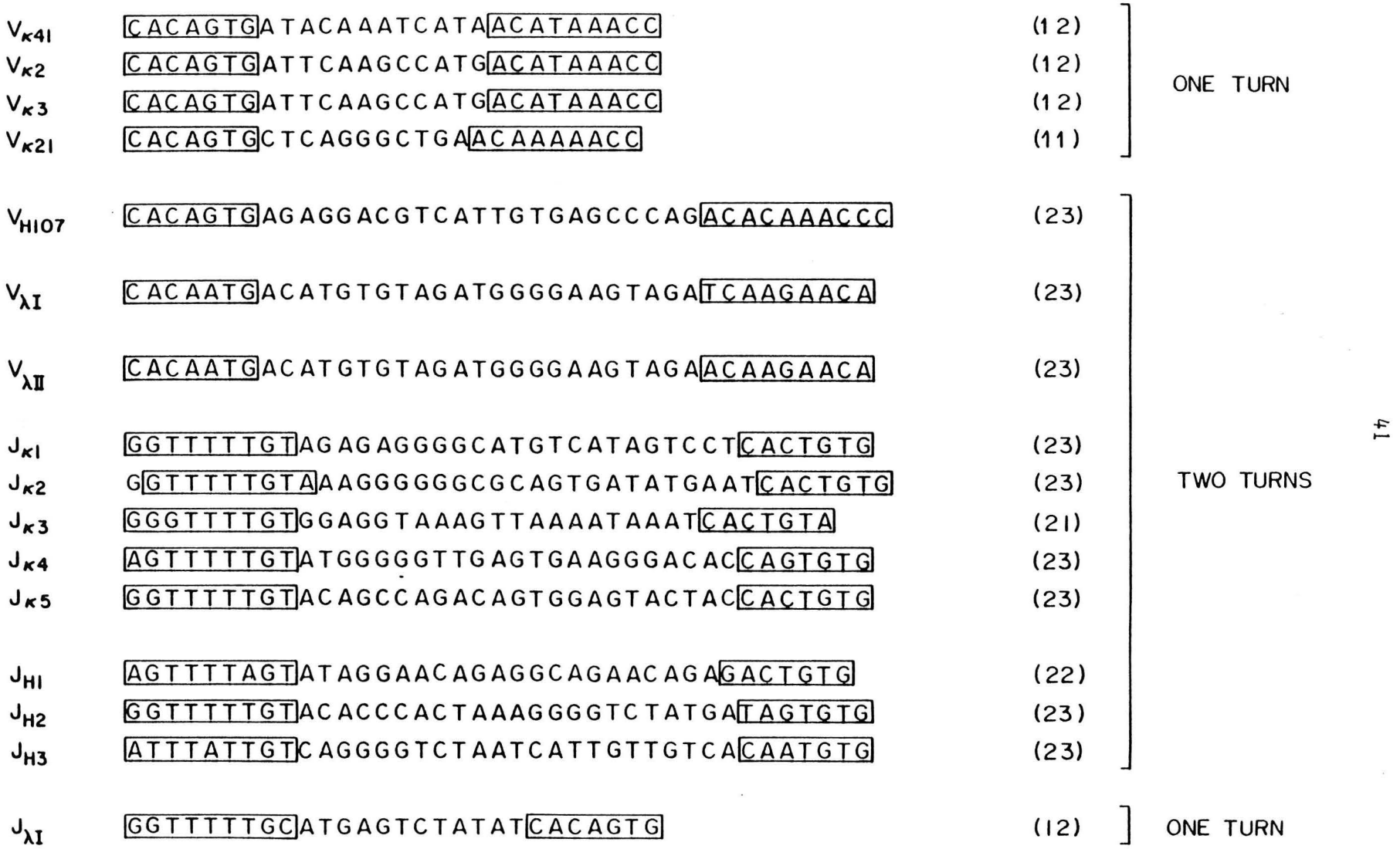

VDJ
Pμ
SμSμSμ
Cμ
Cγ
SγSγSγ
Pα
SαSαSα
Cα
heterodimer formation
site-specific recombination

Figure 5. Recognition signals for DNA rearrangement. A) A schematic drawing. Heptamer and nonamer sequences are denoted with a filled-in box and indicated. The 1 and 2 indicate the one-turn and two-turn spacers, respectively. B) Comparisons of recognition signals of different immunoglobulin gene segments. Heptamers and nonamers are boxed and the length of the spacer distances are indicated in parentheses. The nature of the recognition signals (i.e., one-turn or two-turn) is indicated on the right.


KAPPA
LAMBA
HEAVY
V_κ
V_λ
V_H
7 1 9
7 2 9
7 2 9
D_H
9 1 7 7 1 9
9 2 7 J_κ
9 1 7 J_λ
9 2 7 J_H


$V_{\kappa 41}$ CACAGTG ATACAAATCATA ACATAAACC (12)
$V_{\kappa 2}$ CACAGTG ATTCAAGCCATG ACATAAACC (12)
$V_{\kappa 3}$ CACAGTG ATTCAAGCCATG ACATAAACC (12)
$V_{\kappa 21}$ CACAGTG CTCAGGGCTGA ACAAAAACC (11)
ONE TURN
V_{H107} CACAGTG AGAGGACGTCATTGTGAGCCCAG ACACAAACCC (23)
$V_{\lambda I}$ CACAATG ACATGTGTAGATGGGGAAGTAGA TCAAGAACA (23)
$V_{\lambda II}$ CACAATG ACATGTGTAGATGGGGAAGTAGA ACAAGAACA (23)
$J_{\kappa 1}$ GGTTTTTGT AGAGAGGGGCATGTCATAGTCCT CACTGTG (23)
$J_{\kappa 2}$ GGTTTTTGTA AAGGGGGGCGCAGTGATATGAAT CACTGTG (23)
$J_{\kappa 3}$ GGGTTTTGT GGAGGTAAAGTTAAAATAAAT CACTGTA (21)
$J_{\kappa 4}$ AGTTTTTGT ATGGGGGTTGAGTGAAGGGACAC CAGTGTG (23)
$J_{\kappa 5}$ GGTTTTTGT ACAGCCAGACAGTGGAGTACTAC CACTGTG (23)
J_{H1} AGTTTTAGT ATAGGAACAGAGGCAGAACAGA GACTGTG (22)
J_{H2} GGTTTTTGT ACACCCACTAAAGGGGTCTATGA TAGTGTG (23)
J_{H3} ATTTATTGT CAGGGGTCTAATCATTGTTGTCA CAATGTG (23)
TWO TURNS
$J_{\lambda I}$ GGTTTTTGC ATGAGTCTATAT CACAGTG (12)
ONE TURN

CHAPTER TWO

THE STRUCTURE AND EVOLUTION OF A V_H GENE FAMILY

Portions of this chapter have been submitted for publication to the

Proceedings of the National Academy of Sciences, USA

The Structure and Evolution of a V_H Gene Family

GERALD SIU*, STEPHEN T. CREWS[1], ELIZABETH A. SPRINGER[2],
HENRY V. HUANG[3], and LEROY E. HOOD

**The Division of Chemistry and Chemical Engineering, and the*
Division of Biology, California Institute of Technology
Pasadena, CA 91125, U. S. A.

Present addresses:

[1]Department of Biological Sciences, Stanford University, Stanford, CA 94305.

[2]Northwestern University School of Medicine, Chicago, IL 60611.

[3]Department of Microbiology and Immunology, Washington University School of Medicine, St. Louis, MO 63110.

Abstract

We have sequenced and analyzed the flanking and coding regions of four closely related V_H gene segments referred to as the the T15 V_H gene family. These sequences were also compared to all of the other V_H gene segment sequences that are available. We have identified highly conserved sequences that are partially gene-specific and partially shared as complementary sequences between V_H and V_L promoter regions, implying both a common as well as a chain-specific regulatory element in immunoglobulin expression. The shared sequence is found to be present in one copy in the immunoglobulin enhancer regions as well as histone H2B promoter regions. Analysis of the coding regions of the members of the T15 family indicates that the members of this family duplicated from a common ancestor very recently, and that this V_H gene family is under little selective pressure to maintain its coding sequence.

1. Introduction

The antibody molecule is composed of two polypeptides: a heavy chain and a light chain. These chains consist of a variable region that forms the antigen-binding site, and a constant region that carries out the effector functions, such as complement fixation. The gene that encodes the variable region is composed of several gene segments that are separated in germline DNA and joined together during lymphocyte differentiation (for review, see Early and Hood, 1981). The light chain variable region gene is comprised of V_L and J_L gene segments (Brack *et al.*, 1978; Seidman *et al.*, 1979; Sakano *et al.*, 1979) and the heavy chain variable region gene is composed of V_H, D, and J_H gene segments (Early *et al.*, 1980; Sakano *et al.*, 1980). The V_H gene segment encodes the 100 N-terminal-most amino acids, the D segment encodes the next 1-7 amino acids, and the J_H segment encodes the final 15 amino acids of the variable region (Early *et al.*, 1980). The V_H gene segment family is both extremely large and very diverse; current estimates indicate that the total size of the family could be as large as 100 to 300 members that can differ by as much as 60% in nucleotide sequence (Brodeur and Riblet, in press; G. Siu and S. Crews, unpublished observations). There are a number of important questions concerning the structure of V_H gene segments. It is believed that these genes have duplicated from a common ancestor (Hood *et al.*, 1983); mechanisms for regulation of V_H gene expression presumably have been established for these genes, and may be conserved throughout the V_H gene family. A careful study of the flanking and coding sequences of V_H genes may provide insight into the mechanisms of gene expression in this family.

Unlike most multigene families, the V_H gene segment family is maintained at levels of high diversity. The apparent necessity for size and diversity in this family presents an opportunity to study a unique case in evolution. Many

multigene families, such as histone and rDNA families, are maintained at high-copy number, but at low levels of diversity. This is believed to be a result of gene duplications and frequent gene conversion events (Federoff, 1979; Hentschel and Birnstiel, 1981). Other multigene families, such as globins, are maintained at low-copy number (Fritsch *et al.*, 1980; Lauer *et al.*, 1980). Since the individual members of the globin family have distinct and separate functions and are expressed at different times, they share only weak homology to each other (Lauer *et al.*, 1980; Fritsch *et al.*, 1980; Efstratiadis *et al.*, 1980). In contrast to these other families, the V_H gene segment family is large and diverse. Different V_H gene segments, in combination with different D and J_H segments and light chains, may bind a wide variety of antigens. The spectrum of antigens that one V_H gene segment is capable of binding may overlap with the spectrum of antigens that a different V_H gene segment is capable of binding. It is possible, then, that different V_H gene segments may be employed against the same antigen (Perlmutter *et al.*, submitted); if so, deleterious mutations to one of the V_H gene segments would not necessarily lead to the loss of response to that antigen. Because of this, the selective forces that act on this family are believed to be unique and varied. An analysis of this V_H gene segment family allows a better understanding of the selective pressures that operate on variable region gene segments and the mechanisms that generate diversity.

We have undertaken an extensive analysis of the V_H gene segment family. We have analyzed a group of four closely related V_H gene segments referred to as the T15 V_H gene family (Crews *et al.*, 1981), and we have compared them to all of the other V_H gene segment sequences that have been published to date. In this analysis, we find that the basic structure of the V_H gene segment is only 600-700 base pairs (bp) in length. There are conserved sequences immediately 5' to mRNA transcribed sequences that include sequences that may be necessary for

transcription of the antibody gene, and there are sequences 3' to the gene that are involved in joining the V_H gene segment to the D gene segment. Sequences flanking these conserved sequences do not appear to be conserved among V_H gene segments. Calculations indicate that these gene segments are mutating rapidly, and yet are under some selective pressure to maintain their coding sequence.

2. Materials and Methods

(a) *Sources of cloned DNA and preparation of subclones*

The four gene segments of the T15 V_H gene family, designated V1, V3, V11 and V13, were isolated from a library of BALB/c mouse sperm DNA cloned into a Charon 4A bacteriophage vector (Crews *et al.*, 1981). The clones were isolated by virtue of their hybridization to the plasmid, pS107V1, which was radioactively labeled with ^{32}P by nick translation (Rigby *et al.*, 1977). The pS107V1 plasmid contains a complete copy of the variable region sequence of the mRNA from the S107 myeloma (Early *et al.*, 1980). Its V_H segment sequence is identical to the V1 germline gene segment. Five clones named λV1, λV3, λV11, λV13 and λV19 were used for sequence analysis (Figure 1). Each of the four clones designated its corresponding gene segment, with the exception of λV19, which contains the V1 gene segment. Subclones from each of the λ clones were constructed and transformed into the bacterial host strains HB101 and MC1061. The subclones were constructed as follows:

V1 gene segment

pBV1B2. The 2.4 kb Bam HI fragment from λV1 containing the V1 gene segment was cloned into the Bam HI site of pBR322.

pBV1B5. The 0.54 kb Bam HI fragment from λV1 immediately 5' to the V_H gene was cloned into the Bam HI site of pBR322.

pBV19B4. The 2.4 kb Bam HI fragment from λV19 containing the V1 gene segment

was cloned into the Bam HI site of pBR322. This clone is equivalent to pBV1B2 and the two were used interchangeably.

pBV19R2. The 8.2 kb Eco RI fragment from λV19 containing the V1 gene segment was cloned into the Eco RI site of pBR322.

V3 gene segment

pUV3B3. The 4.4 kb Bam HI fragment from λV3 containing the V3 gene segment was cloned into the Bam HI site of pUK2.

pV3V+. The 1.9 kb Sau 3A fragment containing the V3 gene segment was cloned into the Bam HI site of pUK2.

V11 gene segment

pBV11B3. The 6.2 kb Bam HI fragment from V11 containing the V11 gene segment was cloned into the Bam HI site of pBR322.

pBV11R1. The 5.7 kb Eco RI fragment from V11 containing the V11 gene segment was cloned into the Eco RI site of pBR322.

pBV11Sau. A 1.6 kb Sau 3A fragment from V11 containing the V11 gene segment was cloned into the Bam HI site of pBR322.

V13 gene segment

pBV13H4. The 1.7 kb Hind III fragment from V13 containing the 5' part of the V13 gene segment was cloned into the Hind III site of pBR322.

pBV13H5. The 1.1 kb Hind III fragment from V13 containing the 3' part of the V13 segment was cloned into the Hind III site of pBR322.

(b) *Restriction enzyme mapping and DNA sequencing*

Restriction mapping of the subclones was carried out by either comparing double digests to single digests of various restriction enzymes or by the method of Smith and Birnstiel (1976). The bulk of the DNA sequencing was done using the method of Maxam and Gilbert (Maxam and Gilbert, 1980). Additional sequencing was done utilizing the dideoxynucleotide method of Sanger (Sanger *et al.*, 1977)

utilizing restriction enzyme fragments subcloned into M13mp2 and M13mp8 (Messing and Vieira, 1982). All of the sequence information analyzed were sequenced on both strands in order to minimize errors.

(c) *Analysis of DNA sequences*

DNA sequences were compared by utilizing the dot matrix computer program. In the method used in this paper, the computer compares a five-nucleotide string from one sequence to all five-nucleotide strings in the other sequence; a dot is recorded for each perfect match. Alignment of sequences were performed by inspection. Analysis of mutation rates and divergence times was carried out utilizing the method of Kimura (1981).

3. Results

We have sequenced approximately four kilobases of flanking, coding and intervening sequences 5' and 3' to the four members of the T15 family. The sequencing strategies and restriction maps for each of these genes is shown in Figure 1. Utilizing these sequences and sequences derived from previous work (Crews *et al.*, 1981; Huang *et al.*, 1981), we have analyzed the various features of both the members of the T15 gene family and the V_H gene segment family as a whole using heteroduplex formation, dot matrix analysis, and direct nucleotide comparison.

(a) *General features of the T15 V_H gene family*

V1 gene segment. The V1 gene segment is a functional gene segment; it is utilized in antibodies that bind phosphorylcholine. Its structure is similar to other functional V_H gene segments (Figure 2). The 5' end of the mRNA that encodes the V1 gene expressed in the antiphosphorylcholine response has been identified at a site 63 bp 5' to the initiator codon of the leader peptide (Clarke *et al.*, 1982). Twenty-three base pairs 5' to the initiation site of transcription is a sequence

resembling the TATA sequence that is generally found 22-23 bp 5' to most initiation sites of RNA polymerase II-transcribed genes. The intervening sequence of the V1 gene segment is characterized by a large stretch of alternating Ts and Gs. This stretch of TG dinucleotides is 64 nucleotide pairs long and is not a feature of any other known V_H gene segment.

V3 gene segment. The V3 gene segment is clearly nonfunctional (Huang *et al.*, 1981). Although the 3' half of the sequence encoding the leader peptide bears homology to the other leader sequences in this family, the 5' half does not share homology with the other leader sequences of this family, nor does it share homology with the leaders of any other V_H gene segments. However, there are in-frame ATG initiator codons without an intervening in-frame termination codon, so it is still possible that a leader peptide could be synthesized using this sequence. In addition, the V3 gene segment does not appear to have any of the promoter elements found in other eukaryotic genes and in the other V gene segments (see below). It is possible that the V3 gene segment may have a long 5' untranslated region. As there does not appear to be an adequate promoter in the sequenced region, this would require a 5' untranslated region of over 500 bp.

V11 gene segment. The V11 gene segment is a functional gene segment; it encodes the heavy chain variable region observed in the M47A myeloma antibody (Robinson and Appella, 1979). All of its features are similar to other V_H gene segments.

V13 gene segment. The V13 gene segment is utilized by the 38C myeloma (M. Potter, personal communication). As for V1 and V11, all of its general features are similar to other V_H gene segments.

(b) *Comparisons of the members of the T15 V_H gene family*

We have analyzed the sequences of this family on several levels since different levels of analysis may reveal features that other methods may not.

Heteroduplex analysis of the germline clones of the T15 family (K. Calame, unpublished) reveals that the homology between V11 and V13 extends for long distances in both the 5' and 3' flanking regions. In contrast to the V11-V13 comparison, the homology shared between V3 and V11 and V3 and V13 is very limited; V3 shared only 700 bp of homology with these other gene segments, centered about the V_H coding region.

Figure 3(a-f) show 5 x 5 dot matrices utilizing the flanking and coding sequences of each of the members of the T15 V_H gene family. This method of analysis is useful in that it identifies and characterizes large stretches of homology in sequenced regions and permits the identification of small deletions and small repeated sequences. It is clear from this analysis that the overall levels of homology are decreased in the flanking regions with respect to the coding regions. Of the four members of this family, V11 and V13 appear to be most homologous in the flanking region, although there appears to be a large deletion (84 bp) in the 3' flanking region of V13 with respect to the others. V1 is less homologous in both 5' and 3' flanking regions to V11 and V13, while V3 shares no homology with any of the others in the 5' flanking region immediately preceding the leader sequence. V3 does share homology in the immediate 3' flanking region with the others. The homology in the intervening sequence varies considerably. V11 and V13 have high homology in this region, higher than even the coding region. V3 shares considerably less homology in the intervening sequence with respect to the coding region in comparison with either V11 or V13. V1 shares little or no homology with any of the other members of this family in this region. The leader sequences appear to be very highly conserved, although the small size of the leader makes analysis on this level difficult. The levels of homology are clearly highest in the V_H coding region. In every case (with the exception of the V11-V13 comparison), the homology between members of this family is extremely high in the coding region, higher than any other region.

The sequences were then directly analyzed. Gaps were inserted to maximize homology (Figure 2). In order to learn more about the rates at which these sequences are evolving, we have divided the V_H gene segment region into various regions that may differ functionally and calculated the number of substitutions between each pair of genes. The data were then analyzed using the 3ST model formulae for estimating mutation rates and evolutionary distances proposed by Kimura (1981). Within coding regions we have calculated the mutation rate at each of the three positions constituting a codon and have also calculated the rate for those mutations that cause an amino acid change (replacement-site mutations) and those that do not cause an amino acid change (silent-site mutations). Using the 3ST model, a value for the frequency of base substitutions per site (K) and an error variance for this value (σ_k^2) can be obtained. These values permit us to calculate the rate at which two sequences have diverged from each other as well as an estimate of the error of these values (Figure 4). The mutation frequencies in the flanking regions are much higher than the mutation frequencies in the coding regions. The mutation frequency in the 5' flanking region appears to be the highest in every comparison, significantly higher than even the mutation frequency in the 3' flanking region. The 3' flanking region appears to be mutating somewhat slower than the 5' flanking region; the mutation frequency in each comparison in the 3' flanking region is significantly lower than the mutation frequency in the 5' flanking region. This disparity is somewhat surprising; the 3' flanking region was previously believed to be under no selection pressure and therefore should not differ in mutation frequency from the 5' flanking region. Although the 3' flanking region contains the recognition signals for V-D joining, these sequences are too small to significantly affect the calculated mutation frequencies for this area.

The 5' proximal region, defined as the region starting at the cap site and proceeding 5' to a conserved putative promoter sequence (see below), appears to be mutating at the same rate as the 5' untranslated region. Both of these mutation frequencies are significantly lower than the mutation frequency in the 5' flanking region and somewhat lower than the mutation frequency in the 3' flanking region. The intervening sequences of V3, V11 and V13 appear to be mutating at a rate less than or equal to the mutation rate in the 5' proximal and untranslated regions. The V1 intervening sequence, however, does not appear to be very similar to the other intervening sequences; homology between V1 and the other members of the family in this region is limited almost solely to the splice signals. In addition, the V1 intervening sequence contains a repeating unit of 32 TG nucleotide pairs that is absent in other V_H gene segments. The coding region appears to be mutating the slowest. As expected, the silent substitution frequency is higher than the replacement substitution frequency. The relative selectivity, which we will define as K_A/K_S (the ratio of the mutation frequency of the amino acid replacement-site to the mutation frequency of the silent-site) ranges from 0.45 to 0.8, which is generally higher than most eukaryotic genes. The comparison of the V3 pseudogene to the other functional genes shows that the relative selectivity is higher than the other comparisons (K_A/K_S = 0.6-1.6). This is not unexpected since the V3 gene will be free to diverge at all positions once it is rendered nonfunctional; since there are more possibilities for a replacement substitution than a silent substitution, replacements will accumulate at a high rate. When comparisons are made between the different codon positions, it is apparent that the mutation frequencies in the second and third positions are approximately equal and are consistently higher than the mutation frequency in the first position. Analysis of the mutation frequencies of the different positions of the codon show that for the T15 family, the ranking in order of decreasing value is $K_2>K_3>K_1$; for other eukaryotic genes, the order is $K_3>K_1>K_2$.

4. Discussion

(a) *Structure of V_H gene segments*

The comparative analysis of different V_H gene segments has revealed sequences both 3' and 5' to the coding regions that are shared by all V_H gene segments and may serve a role in the function of these genes.

3' flanking sequences. The putative recognition sequences for V-D joining, a conserved 7 mer and a conserved 9 mer, are located 3' to the V_H gene segment. In addition to these sequences, there appear to be other conserved sequences directly 3' to the V_H gene segment (Figure 5): the first is the dinucleotide sequence CA located 10 bp downstream from the 7 mer and the second is the sequence TGAG located 4 bp downstream from the first sequence. These sequences are comparatively short, thus the significance of their conservation (if any) is unclear. Further analysis reveals that neither the conserved sequence nor its complement is found in the flanking regions of the D and J_H gene segments. A portion of the conserved sequence is an in-frame stop codon which is the first in-frame stop codon 3' to the V_H gene segment (Early *et al.*, 1980).

V_H coding sequence. V_H segments have been divided into framework and hypervariable regions based on analysis of protein sequences (Wu and Kabat, 1970). Since it is clear that somatic mutation can alter the sequence of germline gene segments that are utilized to synthesize antibodies, it is possible that the framework and hypervariable region distinctions do not relate to the diversity of germline gene segments but are a result of somatic mutation and clonal selection of antibody-bearing lymphocytes. Our comparison of germline gene segments clearly indicates that, at least in part, this is not the case; the distinction between the framework and hypervariable regions is evident in the germline gene segment sequence itself. Figure 6 shows an alignment of the translated sequences of a representative of all of the germline V_H gene segment families published to

date. The clustered mutations are highest in the second hypervariable region (residues 50-55). Using the convention of Wu and Kabat (1970), a variability plot of the V_H gene segment sequences in Figure 6 is shown (Figure 7). The hypervariable peaks are clearly observed in the second hypervariable region; since the data analyzed here represent germline V_H gene segments, no somatic mutational or clonal expansion mechanisms have acted upon these sequences. It is interesting to note that the first hypervariable region, located at residues 31-34, does not appear to be in evidence in this comparison. This would indicate that the first hypervariable region is generated via somatic mutational events alone, and is not inherent in the germline, unlike the second hypervariable region. Thus, amino acid residues in the second hypervariable region appear to be mutating faster than the residues in the framework region. This indicates that either negative selection pressure is operating to maintain the framework region, or positive selection is operating to diversify the second hypervariable region.

As mentioned above, the leader sequence appears to be almost as conserved as the coding region. The conservation appears to be limited to the maintenance of the basic structure of the leader sequence; this includes an uninterrupted stretch of neutral or hydrophobic amino acids flanked by charged residues (Silhavy *et al.*, 1983). Comparisons of the intervening sequences of the different V_H gene segments fail to identify any conserved sequences, with the exception of the splice sequences at the boundaries of the intervening sequences.

5' untranslated sequences. These sequences can vary significantly both in length and sequence. The sequences can differ in size from 63 bp (V1) to 29 bp (V_H101), or even longer (Kraig *et al.*, 1983). There is a clear conservation of sequence spanning about 21 bp just 5' to the leader peptide sequence (Figure 8). This sequence homology is not shared by light chain genes or other eukaryotic genes. Their possible function remains unclear, but there exist several

possibilities. It could function as an entry point or to facilitate ribosome binding to the mRNA, perhaps as a way to divert the biosynthetic machinery to the production of heavy chains. Another possibility is that this region plays a role in mRNA stability or folding. To this end, we have examined both variable and constant region sequences for inverted complementary structures that could base-pair with these sequences but have not found any striking candidates. It is of course also possible that these sequences affect the regulation of expression of this gene, either by acting at the transcriptional or at the processing level. It is important to note that although this region appears to be very homologous in most V_H gene segments (including at least four other V_H gene families, data not shown), there are exceptions: V_H101 has this area deleted, and this region in BCL1 and the human V_H genes HG3 and HA2 does not appear to be very homologous to the others (Figure 8).

5' flanking sequences. By analogy to other prokaryotic and eukaryotic genes (Efstratiadis *et al.*, 1980; McKnight, 1982; Hen *et al.*, 1982; Tavernier *et al.*, 1983; Walker *et al.*, 1983; Dierks *et al.*, 1983), it is expected that there are regions in the 5' flanking sequences of V_H gene segments that will be involved in the initiation and regulation of the transcription of antibody genes. We have identified two regions of conserved sequence 5' to the gene that may play a role in the expression of the gene (Figure 8). The first is 23 bp 5' to the initiation site of transcription (the "cap" site). This homology block is similar to the consensus sequence ATAAA$\frac{AA}{TT}$ observed in other genes transcribed by RNA polymerase II 22-23 bp 5' to the initiation site of transcription (the TATA box) (Breathnach and Chambon, 1981). There is no absolute consensus sequence except an AT-rich character. This region is believed to be important for the correct initiation of transcription.

Of more interest, we have found a conserved sequence block 19 bp 5' to the TATA box that appears to be uniquely associated with V_H gene segments. We have designated this homology region the V_H-box, and divided it into two portions; the A portion has the consensus sequence CA$\frac{T}{G}$GAA and is separated by 3 bp from the B portion which has the consensus sequence ATGCAAAT. The CAAAT part of the B sequence is reminiscent of the sequence CCAAT that is found at variable distances from the TATA box in many eukaryotic mRNA-encoding genes (Benoist *et al.*, 1980). It is possible that the TATA box and CCAAT sequences are generally required by eukaryotic genes for accurate and efficient transcription and that additional sequences are required for gene-specific regulation. Such sequences for the heavy chain variable region may be the (A) and part of the (B) regions of the V_H-box. These sequences could interact with regulatory proteins or other macromolecules or may exert their regulatory influence by virtue of direct DNA alterations upon interaction with the internal nuclear milieu. As part of this analysis, we have examined a large number of antibody light chain variable region sequences for the occurrence of this sequence. Since these genes are coexpressed during antibody synthesis, it is conceivable that they might share regulatory sequences. We are able to discern several homologous sequences in light chains (Figure 9). The TATA box is located 21-23 bp to the "cap" site, and a CCAAT sequence is located 43-56 bp 5' to the TATA box. An additional region of homology was identified between the TATA and CCAAT boxes. This conserved sequence follows the consensus sequence TGATTTGCATGT. We refer to this sequence as the V_L-box. It is interesting to note that the core of this sequence is the perfect complement of portion B of the V_H-box, ATGCAAAT. We have analyzed the regions of the heavy chain and the putative light chain enhancer for the presence of this sequence. Both of these regions were found to contain one copy of the V_L-box sequence. In the heavy chain enhancer, the V_L box is found

starting at position 548 and has the sequence TAATTTGCATA (Gillies *et al.*, 1983); in the light chain enhancer, the V_L box is found starting at position 3590 and has the sequence TAATTTAGCAC (Max *et al*, 1981). Statistically, this sequence will occur at random once in every 262 kilobases of DNA. Analysis of other sequenced regions of the immunoglobulin locus reveal that neither the V_L box nor the V_H box nor their complementary sequences are present in these regions. Analysis of the J_κ-C_κ locus reveals that a portion of the V_L box is present 5' to the J_κ gene segments. No copies of the V_H box were found in this region.

A study of the promoter regions of a number of different genes (globins, class I and class II major histocompatibility genes, histones, heat-shock genes, adenovirus, SV40, β-type interferon, thymidine kinase) reveals that the V_H-box is not found in any other promoter region. The V_L-box, however, is found in the promoter region of the histone H2B gene. As shown in Figure 9, histone H2B promoter regions are characterized by a TATA box, preceded in the 5' direction by a variable distance (0-31 bp) by the sequence CTCATTTGCATAC which is preceded at a distance of 19-31 bp by the sequence CCAAT (Harvey *et al.*, 1982). This sequence bears a striking homology to the V_L-box, and the core sequence ATTTGCAT is identical to the core sequence of the V_L-box and is the perfect complement to part B of the V_H-box. This may indicate a coordinate regulation of the antibody and histone H2B gene families, although this still remains unclear. Expression of histone H2B genes is coupled to the cell cycle and DNA synthesis; it is possible that antibody transcription may in fact be coupled to DNA replication as well.

None of these putative promoter sequences has been found in the 5' flanking region of the V3 gene segment. Although it is possible that the V3 gene segment has an extremely long 5' untranslated region, this would be inconsistent

with the data on the other V_H gene segments (Kataoka *et al.*, 1982; Clarke *et al.*, 1982). It is probable that the promoter region has either been deleted, or it was not duplicated during the original genetic event that created the V3 gene segment. Heteroduplex analysis of the λ clones containing the V3 gene segment with the λ clones containing the other members of the T15 family has revealed that the homology between V3 gene segment and the other V_H gene segments is limited to approximately 700 bp immediately surrounding the V gene segment (K. Calame, unpublished). Either two major deletion events occurred, or, more likely, the event that created the V3 gene segment duplicated only 700 bp of sequence around the gene. This event did not duplicate the V_H promoter; therefore the V3 gene segment was probably incapable of being transcribed and hence a pseudogene since its creation. The other crippling mutations were probably obtained subsequent to that time. Analysis of the V3 gene segment in the B10.P strain of *Mus musculus* (R. Perlmutter *et al.*, in preparation) has shown that two of the crippling mutations in the BALB/c V3 gene segment are not found in the B10.P V3 gene segment, yet the promoter region is still absent. This observation is consistent with the theory that the V3 gene segment was a pseudogene since its creation.

(b) *Regulation of antibody gene transcription*

Antibody gene transcription is a complex process modulated at several stages of lymphocyte differentiation. At the pre-B cell stage, the V_H genes are rearranged and synthesized; V_L rearrangement and expression follows in the subsequent stage of B-cell development. In either case, successful rearrangement of the variable region gene is necessary in order to synthesize a functional mRNA; this implies that some of the regulatory elements which control antibody gene expression lie within or near the constant region genes. Recent observations and experiments have supported this view (Rice and Baltimore, 1982; Oi *et al.*, 1983; Gillies *et al.*, 1983; Queen and Baltimore, 1983; Banerji *et al.*, 1983). At the

plasma cell stage (that is, after antigenic challenge), mRNA levels for both heavy and light chains increase dramatically. The regulation of expression of heavy and light chains, then, differs at the early stages of lymphocyte development, yet appears to be coordinated in the terminal steps of development. The control mechanisms for these different levels of expression are unclear. As we have noticed in this paper, there are structures conserved in the 5' flanking regions of both heavy and light chain V gene segments. Part of each structure is heavy or light chain-specific and part is shared between the two chains as a complementary sequence; the shared portion is also present in both heavy chain and light chain enhancer regions. The role of these structures is unclear, yet their strict conservation in both sequence and position with respect to promoter elements strongly implies that they play an important role in the regulation of expression of antibody genes. Efficient and accurate gene transcription require elements operating at the initiation site of transcription in addition to the general chromosomal conformation; since different genes are expressed under different conditions, it is predicted that gene-specific promoter elements exist that determine the specificity of regulation under these different conditions. The V_H and V_L boxes may be the recognition site for the regulating element that enables immunoglobulin genes to be expressed under specific conditions in lymphocytes. The presence of this sequence in both heavy and light chain enhancer regions raises the possibility of an enhancer-promoter sequence-specific interaction to regulate transcription of immunoglobulin genes.

(c) *The evolution of V_H genes and the T15 V_H gene family*

The evolution of variable region genes poses some interesting questions to which we can obtain answers from comparative structural analyses. It is clear that the specificity of antibodies for antigen is determined by the variable region sequence. Since the ability to synthesize antibodies that can react with certain

pathogenic organisms and harmful substances clearly would be selected for in a positive manner, it is possible that selection may be acting on the primary sequence of the antibody variable region gene. It is, however, difficult to know exactly what the selective pressures acting on a given variable region gene may be since a given heavy or light chain gene segment can be utilized to form antibodies against a number of different substances. In studying the significance of any particular allele in the gene pool, it is difficult to evaluate its selective advantage in a natural population. Variable region gene segments pose an additional problem in that if one gene segment cannot be utilized to make antibodies against a particular antigen, other genes may be utilized to make an effective response. In that sense, the multiplicity of germline V_H gene segments may serve to provide a back-up system to neutralize foreign substances in the event of a damaging mutation to a utilized V_H segment. This multispecificity may imply that a given V_H gene segment may be selectively neutral and only several or a series of genes would be under selective pressure. Alternatively, each gene segment may be considerably better adapted to its role such that it is under direct selective pressure. To answer these questions, we have analyzed the coding region and flanking region sequences of the T15 family (Figure 4).

Analysis of the T15 family indicates that the 5' flanking region is evolving the fastest. The mutation frequency in the 5' flanking region is significantly higher than the other mutation frequencies, even higher than the mutation frequency in the 3' flanking region. This implies that the 5' flanking region is under very little selection pressure to maintain its sequence. This would indicate that the far 5' flanking region distal to the V_H box is not important for either the structure or the function of the immunoglobulin gene and is probably mutating at the high mutation rate of unselected DNA. The 3' flanking region appears to be mutating much slower than the 5' flanking region; as mentioned above, the

mutation frequency in each comparison in the 3' flanking region is significantly lower than the mutation frequency in the 5' flanking region. Although the 3' flanking region contains the recognition signals for V-J joining, these sequences are too small to substantially alter the calculated mutation frequencies for this area. This implies that the 3' flanking region may have an additional function in the proper expression or the structure of the immunoglobulin gene. It has been hypothesized that certain V gene segments are preferentially rearranged to certain D or J segments. If this is the case, one might expect that the control region for rearrangement preference could be located near the recognition signals and the joining site. The lower mutation rate in the 3' flanking region, then, would be a reflection of the selection pressure to maintain this control region.

The 5' proximal region appears to be mutating at the same rate as the 5' untranslated region. Both of these mutation frequencies are somewhat lower than the mutation frequency in the 3' flanking region. The 5' proximal region contains the signals that are believed to be important in the correct expression of the immunoglobulin gene; the lower mutation frequency in this region probably reflects the selection forces that are operating to maintain these control sequences. The 5' untranslated region has no known function, but may be constrained by forces that affect mRNA structure, stability, or function (Shine and Dalgarno, 1974; Kozak, 1981).

The intervening sequence appears to be mutating at a rate slightly slower than the 5' untranslated regions. As this calculation ignores the existence of insertions and deletions, this difference is probably not significant. This implies that there is some negative selection pressure that acts to maintain all mRNA sequences, whether coding or not. As mentioned above, this negative selection pressure may be involved in the maintainence of the mRNA structure or stability. This theory was initially proposed to explain observations that indicated that the

silent-site mutation rate was slower than the mutation rate of unselected DNA (Kafatos *et al.*, 1977). The only highly conserved regions observed are the consensus splice signals (Breathnach and Chambon, 1981) at the boundaries of the intervening sequence.

The leader sequence is evolving at a fairly slow rate, slower than the rates in the intervening sequence and in the 5' untranslated region. The basic structure of a leader sequence is conserved; a string of neutral or hydrophobic amino acids flanked by charged residues at both ends. The slower mutation rate in this sequence probably reflects the negative selection forces which maintain this structure.

We have examined the rate of silent substitutions occurring in the V_H coding sequence of the functional genes of the T15 family and compared them to the rate of substitution of the flanking and intervening sequences. We find that the silent substitution rate is lower than the flanking and intervening sequence rates; this is consistent with data derived from study of other genes (Miyata *et al.*, 1980). This indicates that slight selective pressure does exist to maintain the silent-site sequence. The constraints on nucleotide selection at silent sites is unclear but may have to do with tRNA abundances, anticodon stabilities, or be involved in mRNA stability or secondary structure (Kafatos *et al.*, 1977). When the number of silent-site mutations is compared to the number of replacement-site mutations, we find that the silent mutation rate is 1.5-2.5 times greater than the replacement mutation rate. This implies that there is selective pressure operating to maintain the V_H coding sequence above that which partially maintains the silent-site sequence. This conflicts with the argument stated above that individual V_H gene segments are under little selective pressure since other V_H gene segments which can bind the same antigen can provide an adequate immune response to make up for the potential loss from deleterious mutations. If

this were the case, we would expect the replacement rate to be more comparable to the silent rate. Since this is not the case for any of the members of this family, we believe that many V_H gene segments are under selective pressure to maintain their coding sequence. A study of the T15 V_H gene family in a different inbred strain of *Mus musculus*, B10.P, has been completed (Perlmutter *et al.*, in preparation) and provides evidence consistent with this view. This analysis reveals that the coding region has been conserved at a significantly higher level than the flanking or intervening sequences. This pressure may not be identical throughout the V_H coding sequence; as mentioned above, it has been noted that the hypervariable regions of closely-related germline V_H gene segments tend to have more fixed mutations than the framework regions.

Although the mutation frequencies in the replacement site are lower than the silent site frequencies, they are extremely high. The mutation rates derived from these mutation frequencies are as much as an order of magnitude higher than mutation rates in other genes (Miyata *et al.*, 1980). The relative selectivity (K_A/K_S) is also higher than what is observed for other genes (Miyata *et al.*, 1980). This implies that although the members of this family are under selection pressure, the duplicated members of this family are nonetheless diverging from each other very rapidly. This is further indicated by the pattern of the mutation frequencies in the three positions of the codon. As mentioned above, the pattern of mutation frequencies ranked in order of decreasing value is $K_3>K_1>K_2$. For most eukaryotic genes, the frequency in the second position is generally lowest due to the lack of degeneracy; as there are no degenerate positions in the second position of a codon, a mutation in the second position would guarantee that an amino acid replacement would occur. Furthermore, it has been observed that mutations in the second position are more likely to produce a nonconservative replacement; that is, amino acid replacement changes in the second position result

in greater changes in the physical properties of the protein (Salser, 1976). The mutation frequency in the third position is generally very high due to the increased degeneracy in that position. The high mutation frequency in the second position in the T15 family indicates that the T15 family, and probably the V_H gene family in general, is less restricted by selective forces than other eukaryotic genes and can therefore tolerate more changes in the primary protein structure without loss of function.

Our evolutionary analysis of the T15 family reveals that since the duplication of the ancestral gene that created this family, the members have diverged from each other at an extremely high rate which appears to be characteristic of V gene segments. Despite this rapid mutation rate, there is strong evidence that selection is occurring to maintain at least one of the members of this family.

The authors would like to thank Karyl Minard for her valuable assistance, Sasha Kamb and Drs. Mitchell Kronenberg and Roger Perlmutter for criticisms of the manuscript and helpful discussions, and Connie Katz for help in the preparation of the manuscript.

REFERENCES

Altenburger, W., Steinmetz, M. & Zachau, H. (1980). *Nature* **287,** 603-607.

Banerji, J., Olson, L. & Schaffner, W. (1983). *Cell* **33,** 729-740.

Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). *Nucl. Acids Res.* **8,** 127-142.

Bentley, D. & Rabbitts, T. (1980). *Nature* **288,** 730-733.

Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. (1981). *Cell* **24,** 625-637.

Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. (1978). *Cell* **15,** 1-14.

Breathnach, R. & Chambom, P. (1981). *Ann. Rev. Biochem.* **50,** 349-383.

Brodeur, P. & Riblet, P. (1984). *Eur. J. Immunol.*, in press.

Clarke, C., Berenson, J., Goverman, J., Boyer, P., Crews, S., Siu, G. & Calame, K. (1982). *Nucl. Acids Res.* **10,** 7731-7749.

Cohen, J., Effron, K., Rechavi, G., Ben-Neriah, Y., Zakut, R. & Givol, D. (1982). *Nucl. Acids Res.* **10,** 3353-3370.

Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. (1981). *Cell* **25,** 59-66.

Dierks, P., van Ooyen, A., Cochran, M., Dobkin, C., Reiser, J. & Weissman, C. (1983). *Cell* **32,** 695-706.

Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980). *Cell* **19,** 981-992.

Early, P. & Hood, L. (1981). *Genetic Engineering* (J. K. Setlow and Hollaender, eds.), pp. 157-188, Plenum Press, New York.

Early, P., Notterberg, C., Weissman, I. & Hood, L. (1982). *Molec. Cell. Biol.* **2,** 829-836.

Efstratiadis, A., Posakony, J., Maniatis, T., Lawn, R., O'Connell, C., Spritz, R., DeRiel, J., Forget, B., Weissman, S., Slightom, J., Blechl, A., Smithies, O., Baralle, F., Shoulders, C. & Proudfoot, N. (1980). *Cell* **21,** 653-668.

Federoff, N. (1979). *Cell* **16,** 697-710.

Fritsch, E., Lawn, R. & Maniatis, T. (1980). *Cell* **19,** 959-972.

Gillies, S., Morrison, S., Oi, V. & Tonegawa, S. (1983). *Cell* **33,** 717-728.

Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. & Cohen, J. (1981). *Nature* **292,** 426-430.

Harvey, R. P., Robins, A. & Wells, J. (1982). *Nucl. Acids Res.* **10,** 7852-7863.

Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M. & Chambon, P. (1982). *Proc. Natl. Acad. Sci.* **79,** 7132-7136.

Hentschel, C. & Birnstiel, M. (1981). *Cell* **25,** 301-313.

Hood, L., Hunkapiller, T. & Kraig, E. (1983). *Modern Cell Biology* (J. R. McIntosh, ed.), pp. 305-328, Alan R. Liss, New York.

Huang, H., Crews, S. & Hood, L. (1981). *J. Molec. Appl. Genet.* **1,** 93-101.

Jukes T. (1980). *Science* **210,** 973-978.

Kafatos, F., Efstratiadis, A., Forget, B. & Weissman, S. (1977). *Proc. Natl. Acad. Sci.* **74,** 5618-5622.

Kataoka, T., Niakido, T., Miyata, T., Moriwaki, K. & Honjo, T. (1982). *J. Biol. Chem.* **257,** 277-285.

Kelley, D., Coleclough, C. & Perry, R. (1982). *Cell* **29,** 681-689.

Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. (1981). *Cell* **27,** 573-581.

Kimura, M. (1981). *Proc. Natl. Acad. Sci.* **78,** 454-458.

Knapp, M., Liu, C.-P., Newell, N., Tucker, P., Ward, R., Strober, S. & Blattner, F. (1982). *Proc. Natl. Acad. Sci.* **79,** 2996-3000.

Kozak, M. (1981). *Nucl. Acids Res.* **9,** 5233-5252.

Kraig, E., Kronenberg, M., Kapp, J., Pierce, C., Abruzzini, A., Sorensen, C., Samuelson, L., Schwartz, R. & Hood, L. (1983). *J. Exp. Med.* **158,** 192-209.

Lauer, J., Shen, C.-K. J. & Maniatis, T. (1980). *Cell* **20,** 119-130.

Litman, G. W., Berger, L., Murphy, K., Litman, R., Hinds, K., Jahn, C. & Erickson, B. (1983). *Nature* **303,** 349-352.

Matthyssens, G. & Rabbits, T. H. (1980). *Proc. Natl. Acad. Sci.* **77**, 6561-6565.

Max, E., Seidman, J., Miller, H. & Leder, P. (1980). *Cell* **21**, 793-799.

Max, E., Maizel, H. & Leder, P. (1981). *J. Biol. Chem.* **256**, 5116-5120.

Maxam, A. & Gilbert, W. (1980). *Meth. Enzymol.* **65**, 499-560.

McKnight, S. (1982). *Cell* **31**, 355-365.

Messing, J. & Vieira, J. (1982). *Gene* **19**, 269-276.

Miyata, T., Yasunaga, T. & Nishida, T. (1980). *Proc. Natl. Acad. Sci.* **77**, 7328-7332.

Nishioka, Y. & Leder, P. (1980). *J. Biol. Chem.* **255**, 3691-3694.

Oi, V., Morrison, S., Herzenberg, L. & Berg, P. (1983). *Proc. Natl. Acad. Sci.* **80**, 825-829.

Ollo, R., Auffray, C., Sikorav, J.-L. & Rougeon, F. (1981). *Nucl. Acids Res.* **9**, 4099-4109.

Perlmutter, R., Klotz, J. L., Bond, M. W., Nahm, M., Davie, J. M. & Hood, L. (1983). *J. Exp. Med.*, submitted.

Queen, C. & Baltimore, D. (1983). *Cell* **33**, 741-748.

Rabbitts, T. H., Matthyssens, G. & Hanlyn, P. H. (1980). *Nature* **284**, 238-243.

Rechavi, G., Ram, D., Glazer, L., Zakut, R. & Givol, D. (1983). *Proc. Natl. Acad. Sci.* **80**, 855-859.

Rice, D. & Baltimore, D. (1982). *Proc. Natl. Acad. Sci.* **79**, 7862-7865.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**:237-251.

Robinson, E. & Appella, E. (1979). *J. Biol. Chem.* **254**: 11418-11430.

Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. (1979). *Nature* **280**, 288-294.

Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980). *Nature* **286**, 676-683.

Salser, W. (1976). *Cold Spring Harbor Symp. Quant. Biol.* **62**, 985-1002.

Sanger, F., Nicklen, S. & Coulson, A. (1977). *Proc. Natl. Acad. Sci.* **74,** 5463-5467.

Seidman, J., Max, E. & Leder, P. (1979). *Nature* **280,** 370-375.

Seidman, J. & Leder, P. (1980). *Nature* **286,** 779-783.

Selsing, E. & Storb, U. (1981). *Cell* **25,** 47-58.

Shine, J. & Dalgarno, L. (1974). *Proc. Natl. Acad. Sci.* **71,** 1342-1346.

Silhavy, T., Benson, S. & Emr, S. (1983). *Microbiol. Rev.* **47,** 313-344.

Smith, H. & Birnstiel, M. (1976). *Nucl. Acids Res.* **3,** 2387-2398.

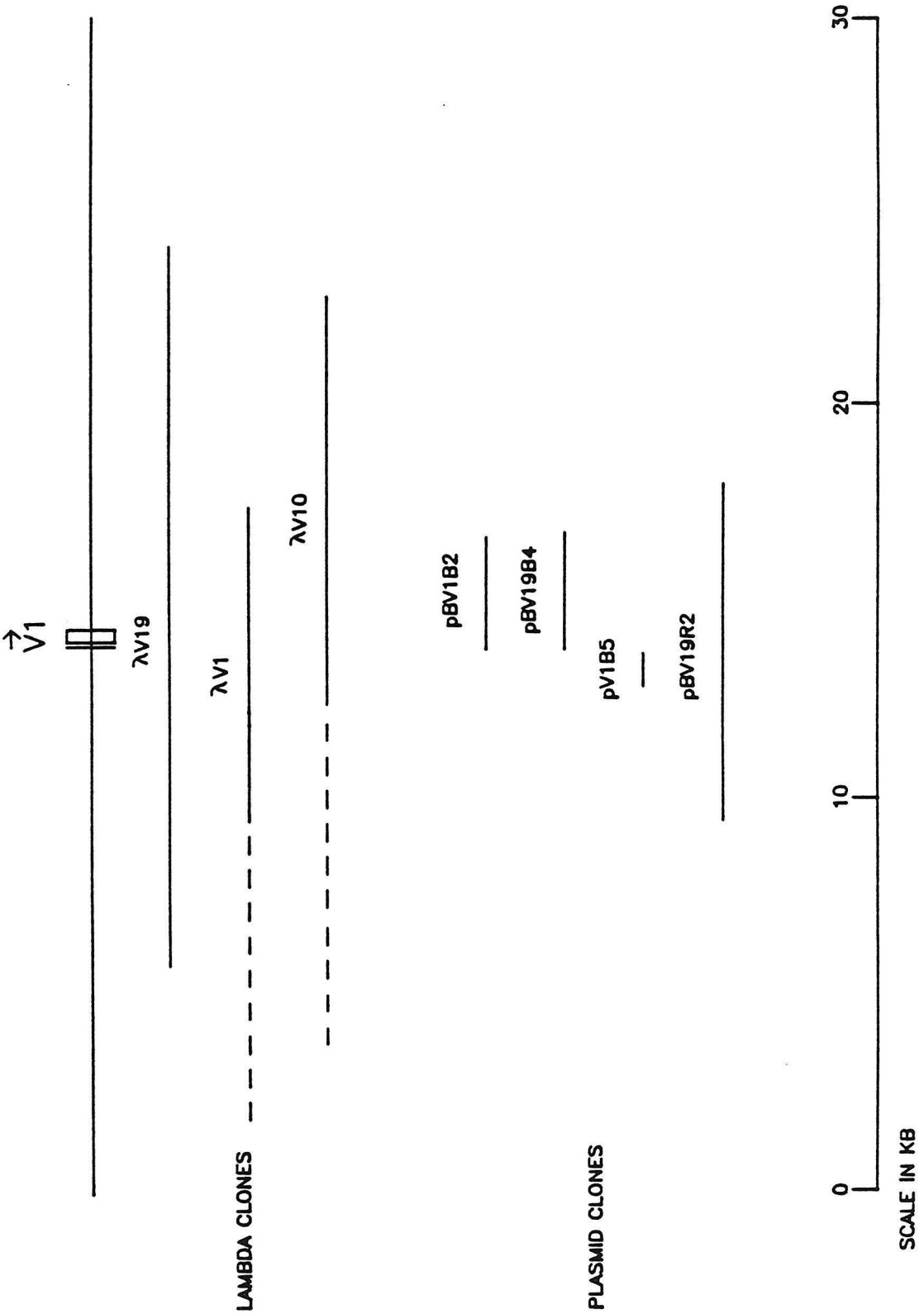
Tavernier, J., Gheysen, D., Duerinck, F., van der Heyden, J. & Fiers, W. (1983). *Nature* **301,** 634-636.

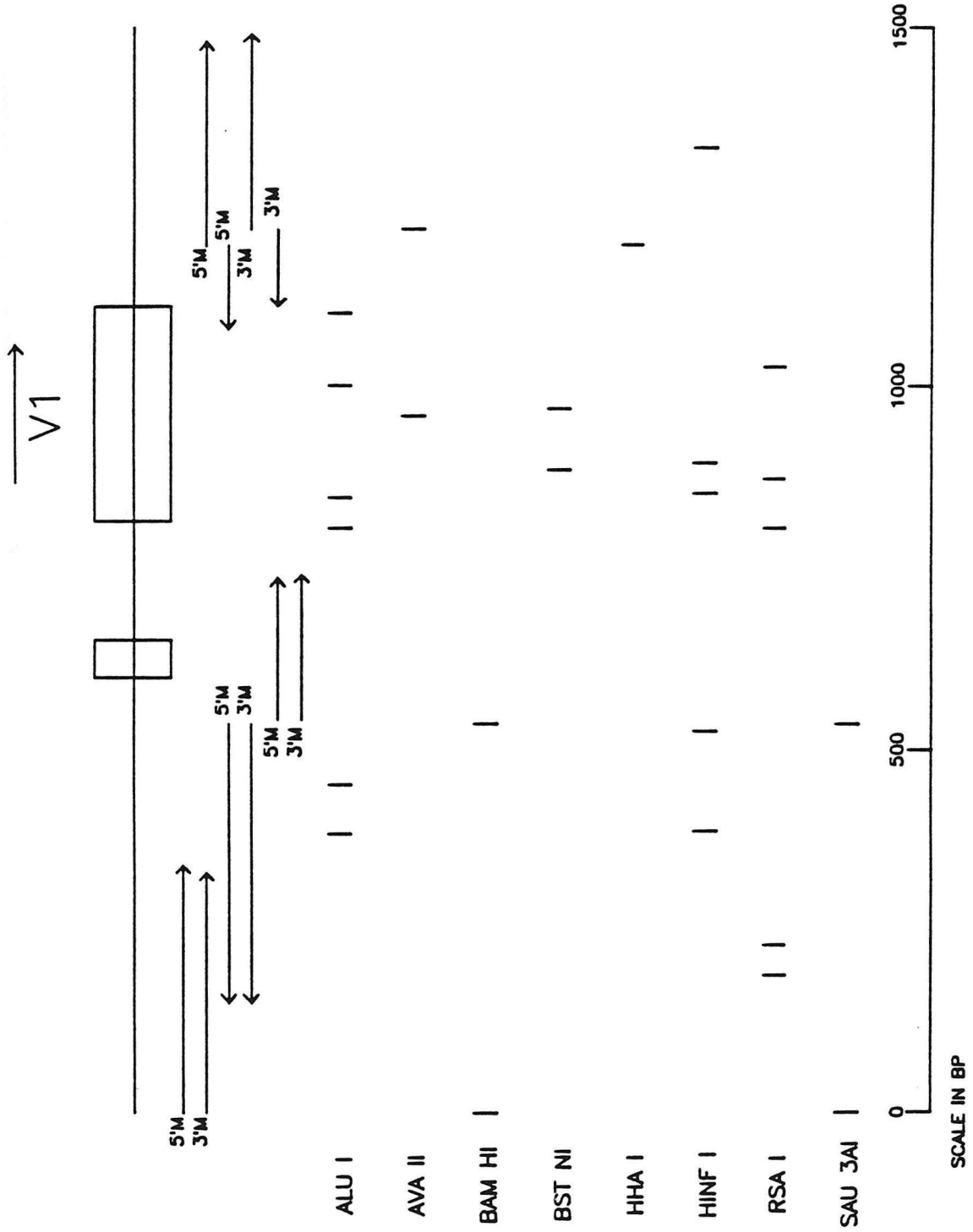
Walker, M., Edlund, T., Boulet, A. & Rutter, W. (1983). *Nature* **306,** 557-561.

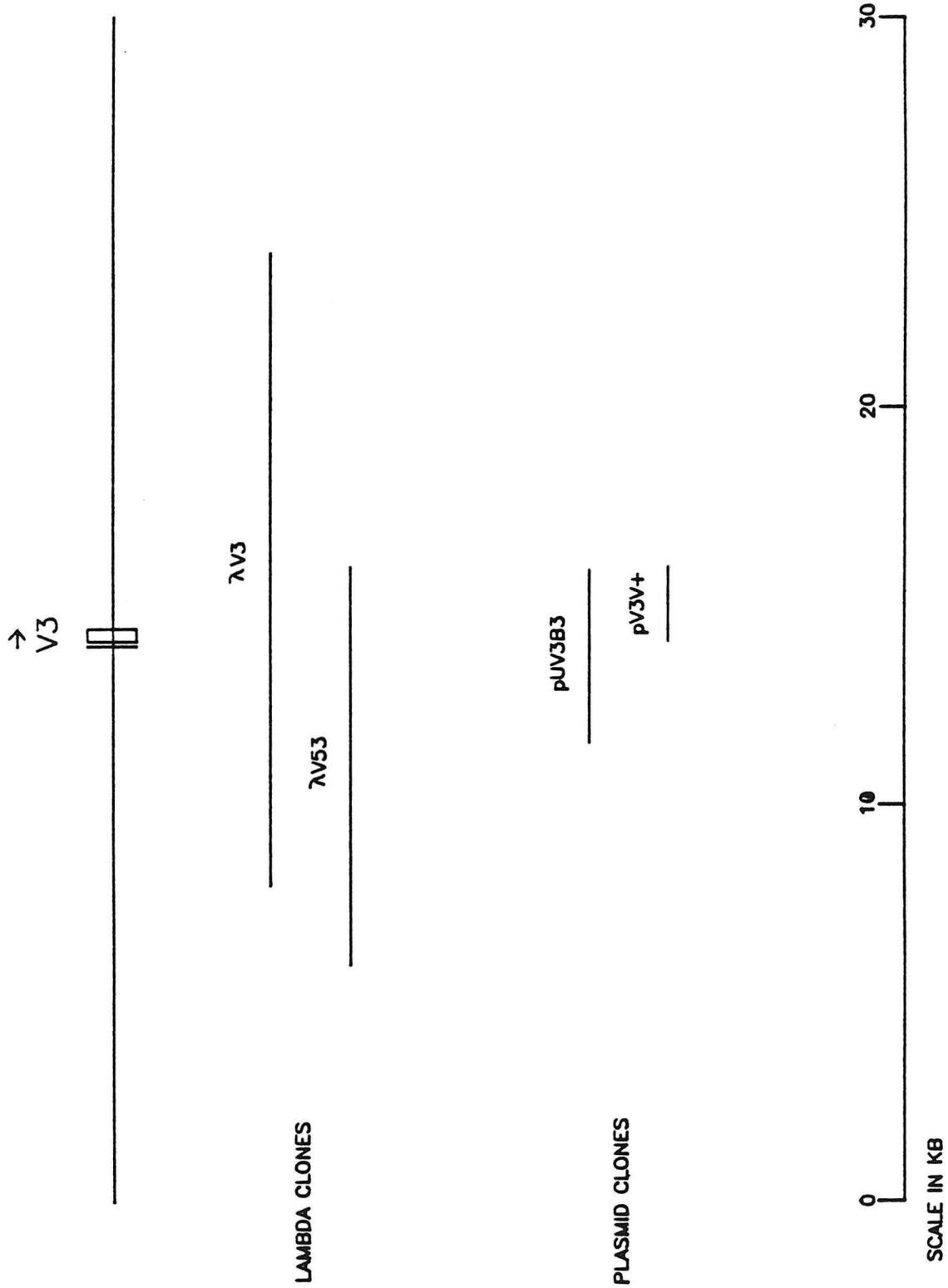
Wu, G., Govindji, N., Hozumi, N. & Murialdo, H. (1982). *Nucl. Acids Res.* **10,** 3831-3843.

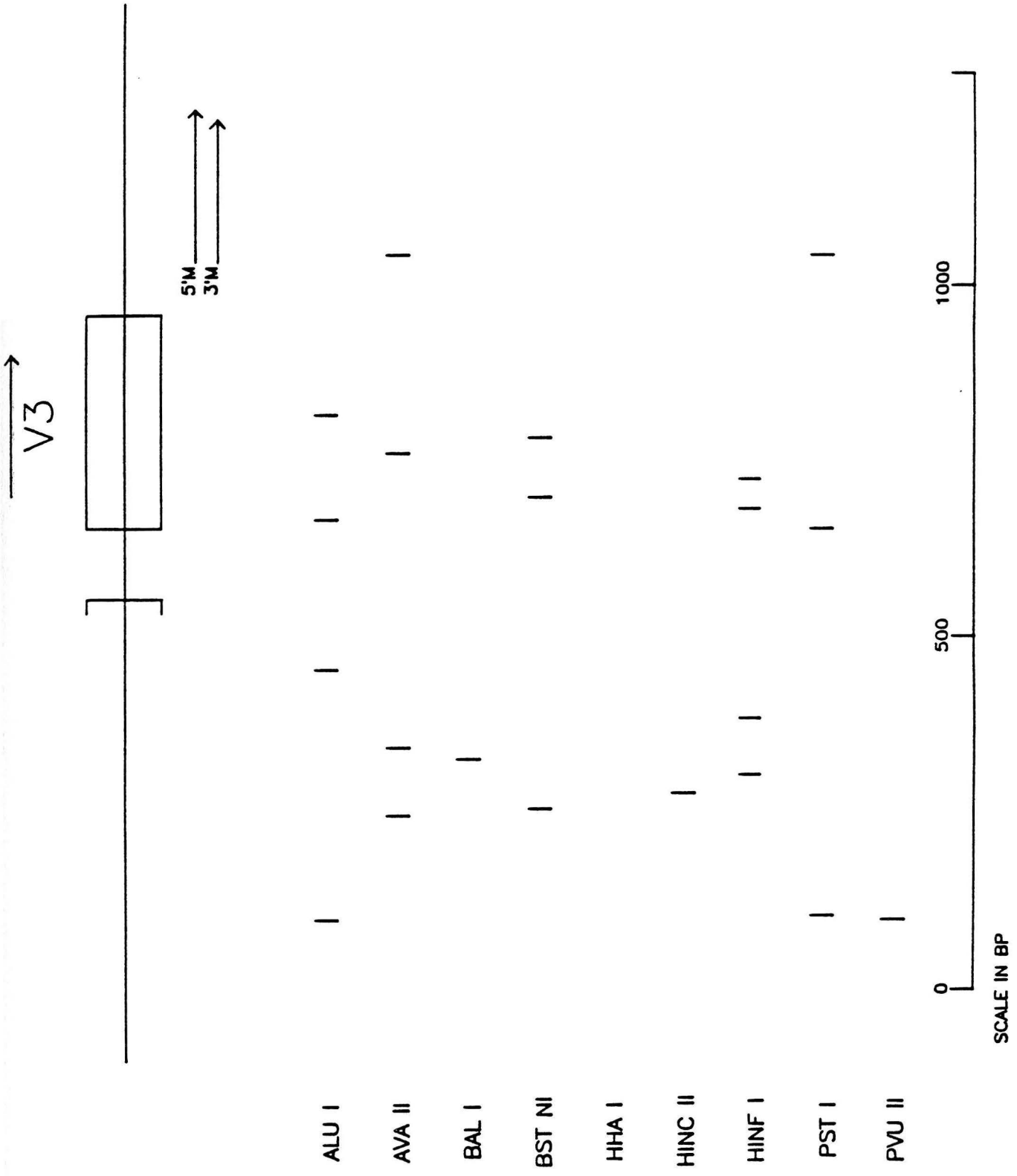
Wu, T. T. & Kabat, E. A. (1970). *J. Exp. Med.* **132,** 211.250.

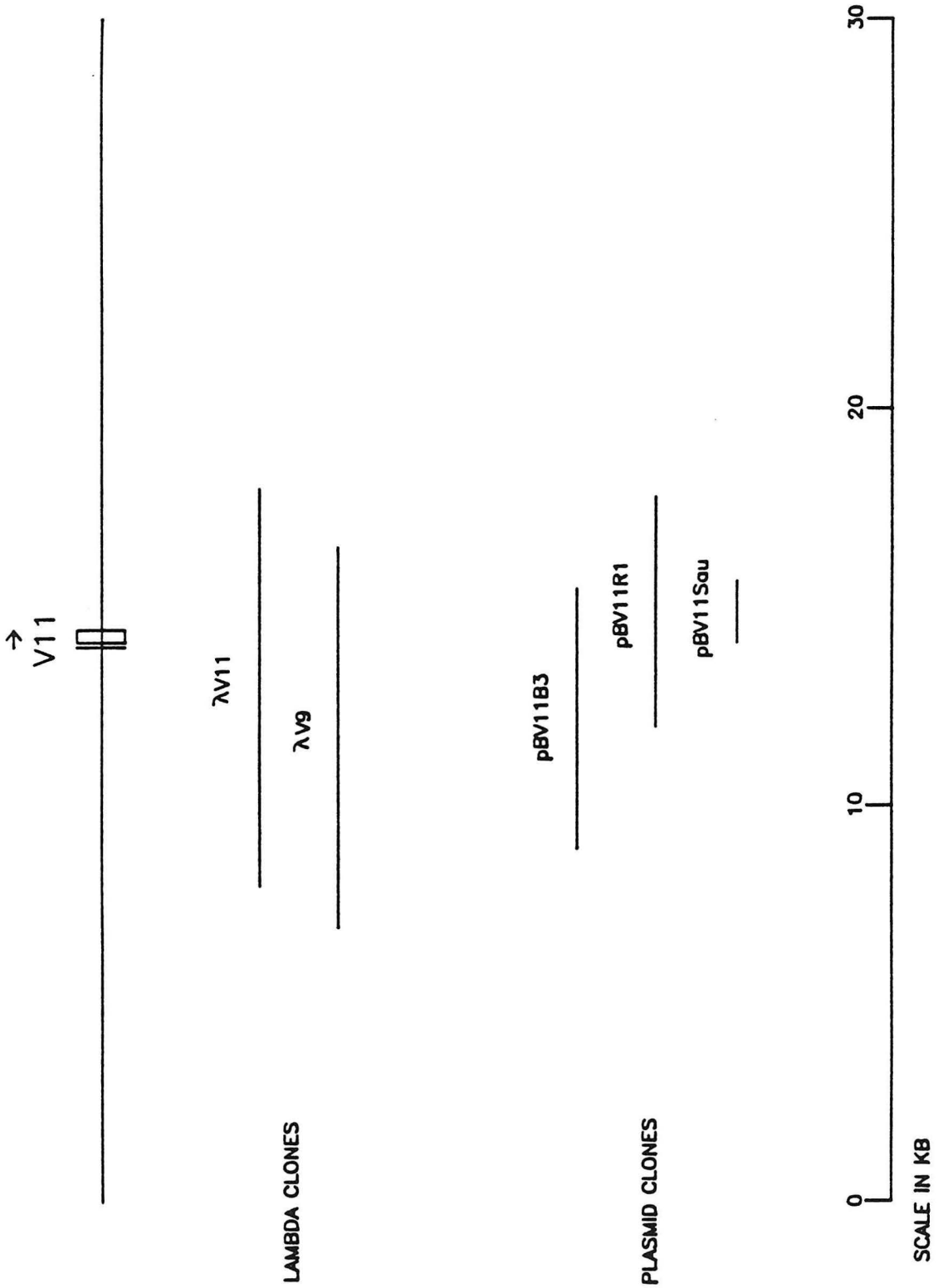
FIG. 1. Bacteriophage lambda genomic clones and plasmid subclones containing the members of the T15 V_H gene family and restriction maps and nucleotide sequencing strategies. a) V1, b) V3, c) V11, d) V13. V_H gene segments are indicated by the open box; the leader exon is indicated by a line. Arrows over the gene segments indicate direction of transcription. Dotted lines in the λV1 and λV10 clones indicate cloning artifacts consisting of the cloning of the Charon 4A "stuffer" fragment next to mouse genomic DNA. Arrows below the gene segments and flanking regions indicate sequencing direction. Designation next to the arrows indicates polarity of end-labeling and technique: M indicates Maxam-Gilbert techniques, S indicates Sanger (dideoxy) techniques. Subsequent sequence analyses of these regions have completed sequence through all of the restriction sites used for sequencing in the previous analyses, with the exception of the Bam HI site in the V1 gene cluster.

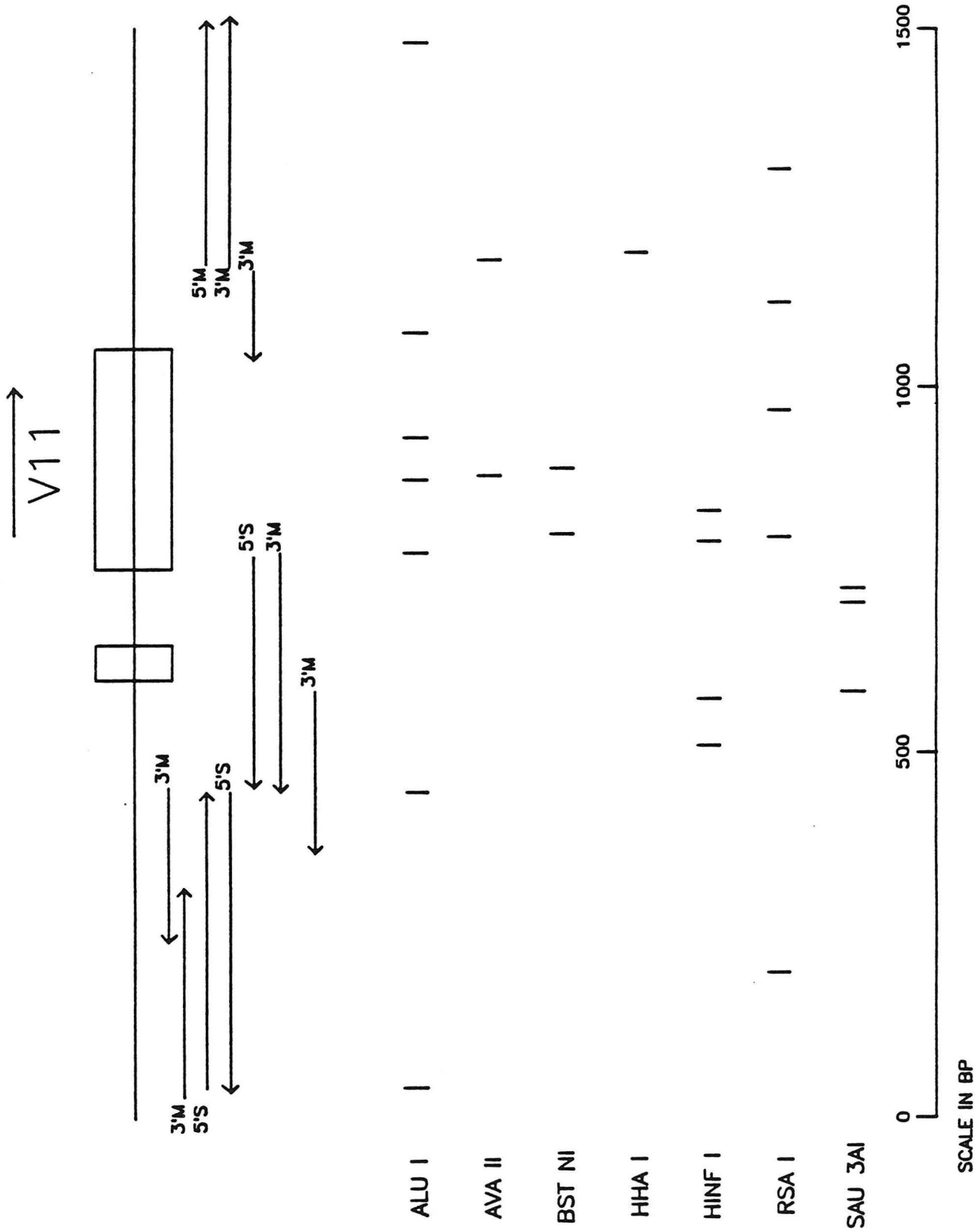

V1
λV19
λV1
λV10
LAMBDA CLONES
pBV1B2
pBV19B4
pV1B5
pBV19R2
PLASMID CLONES
0
10
20
30
SCALE IN KB

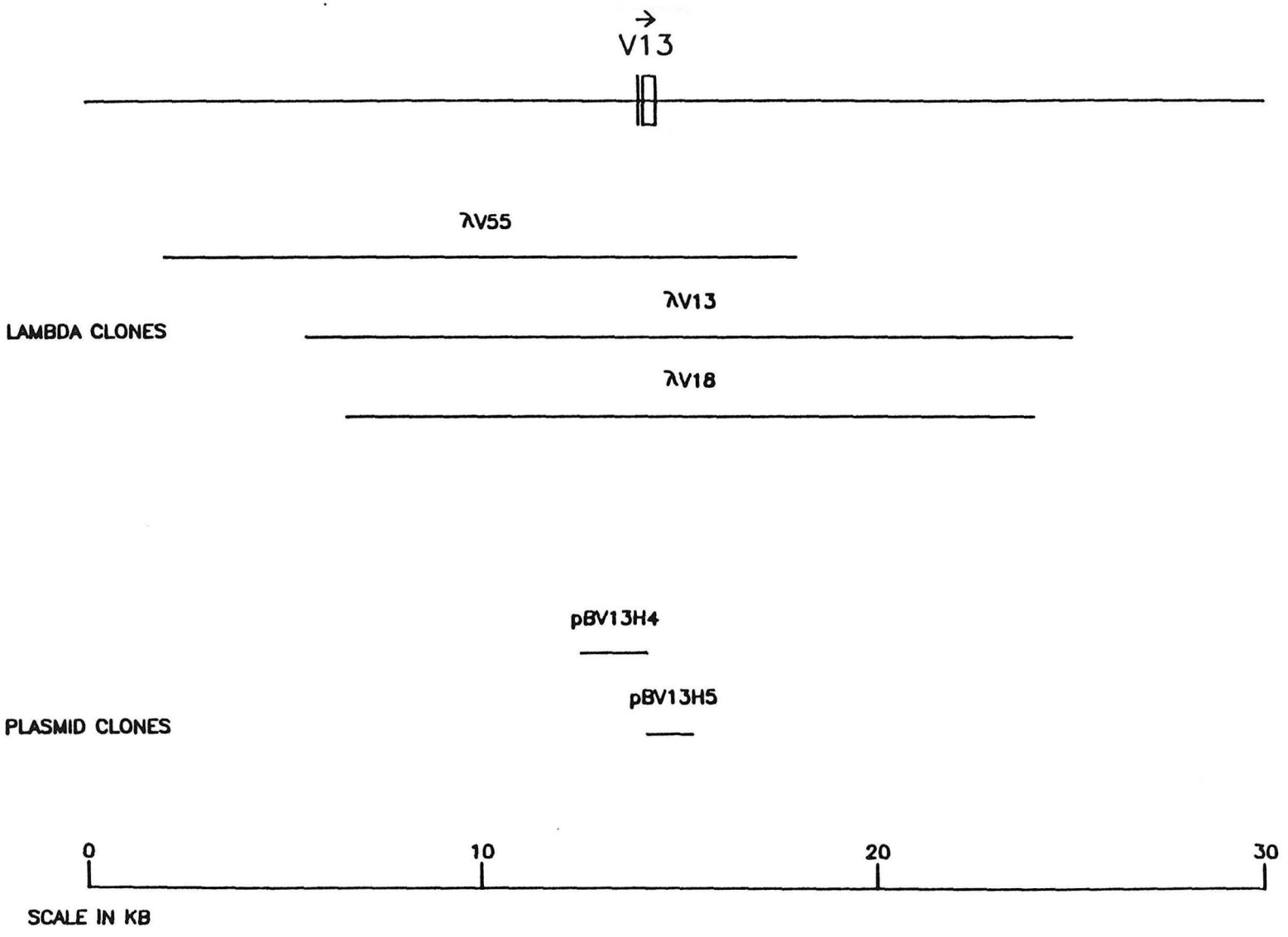

V1
5'M
3'M
5'M
3'M
5'M
3'M
5'M
3'M
5'M
3'M
ALU I
AVA II
BAM HI
BST NI
HHA I
HINF I
RSA I
SAU 3AI
0
500
1000
1500
SCALE IN BP


→
V3
λV3
λV53
LAMBDA CLONES
pUV3B3
pV3V+
PLASMID CLONES
0
10
20
30
SCALE IN KB


V3
5'M
3'M
ALU I
AVA II
BAL I
BST NI
HHA I
HINC II
HINF I
PST I
PVU II
0
500
1000
SCALE IN BP


V11
λV11
λV9
LAMBDA CLONES
pBV11B3
pBV11R1
pBV11Sau
PLASMID CLONES
0
10
20
30
SCALE IN KB


V11
5'M
3'M
3'M
5'S
3'M
3'M
5'S
3'M
3'M
5'S
ALU I
AVA II
BST NI
HHA I
HINF I
RSA I
SAU 3AI
0
500
1000
1500
SCALE IN BP


V13
λV55
λV13
LAMBDA CLONES
λV18
pBV13H4
pBV13H5
PLASMID CLONES
0
10
20
30
SCALE IN KB

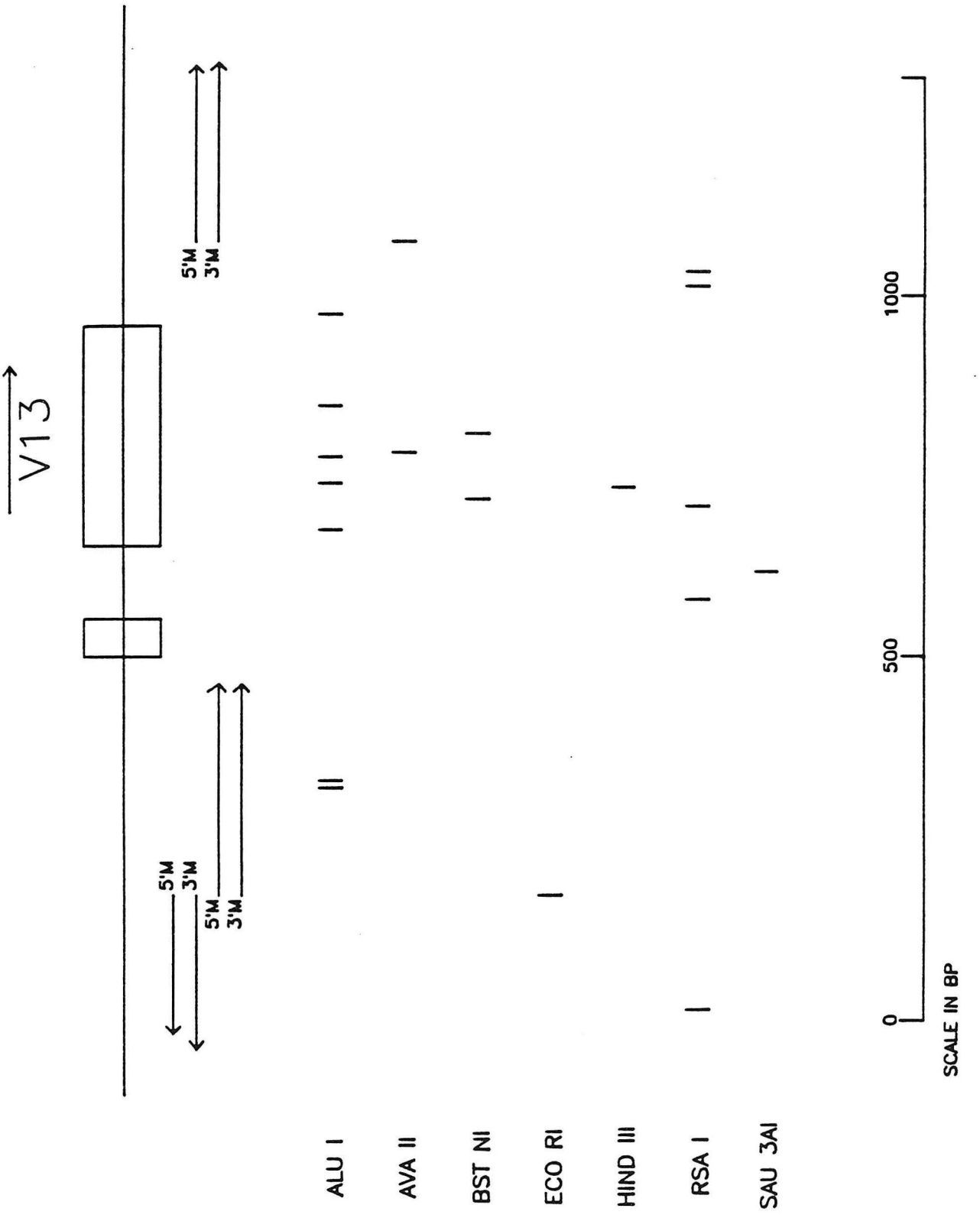

V13
5'M
3'M
5'M
3'M
5'M
3'M
ALU I
AVA II
BST NI
ECO RI
HIND III
RSA I
SAU 3AI
0
500
1000
SCALE IN BP

FIG. 2. Nucleotide sequences of the members of the T15 family and their flanking regions. The cap site for the RNA transcribed from the V1 gene has been identified (Clarke *et al.*, 1982) (boxed "A" at position 575). The leader sequence and the coding regions are boxed; the TATA homology and the rearrangement recognition signals are indicated. Gaps were inserted to maximize homology.

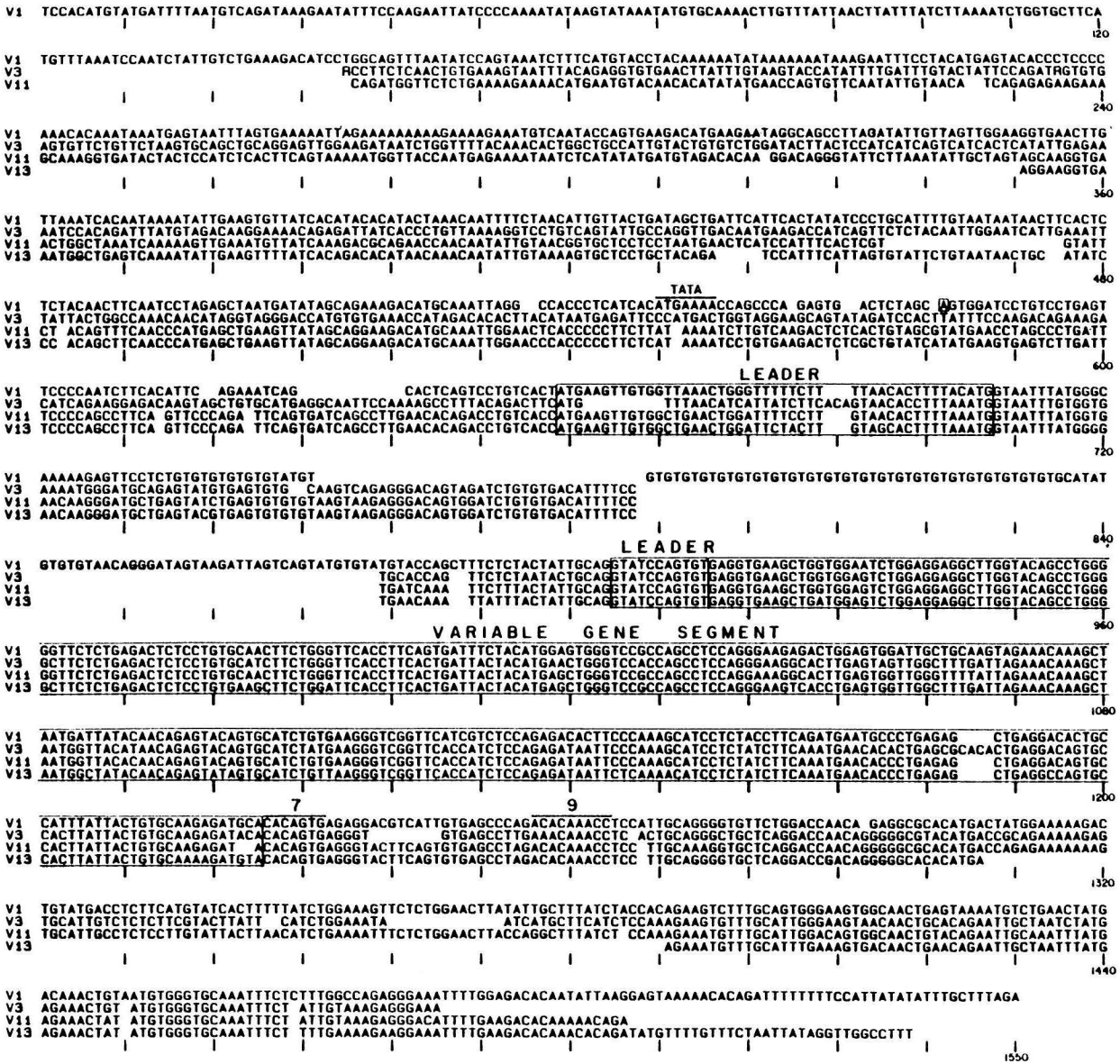
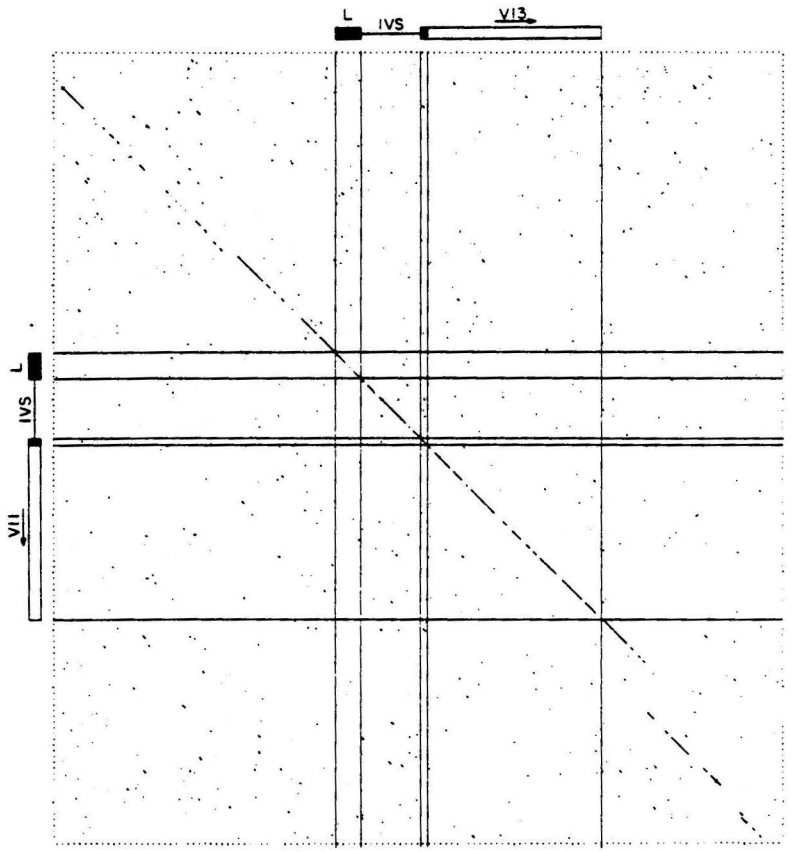
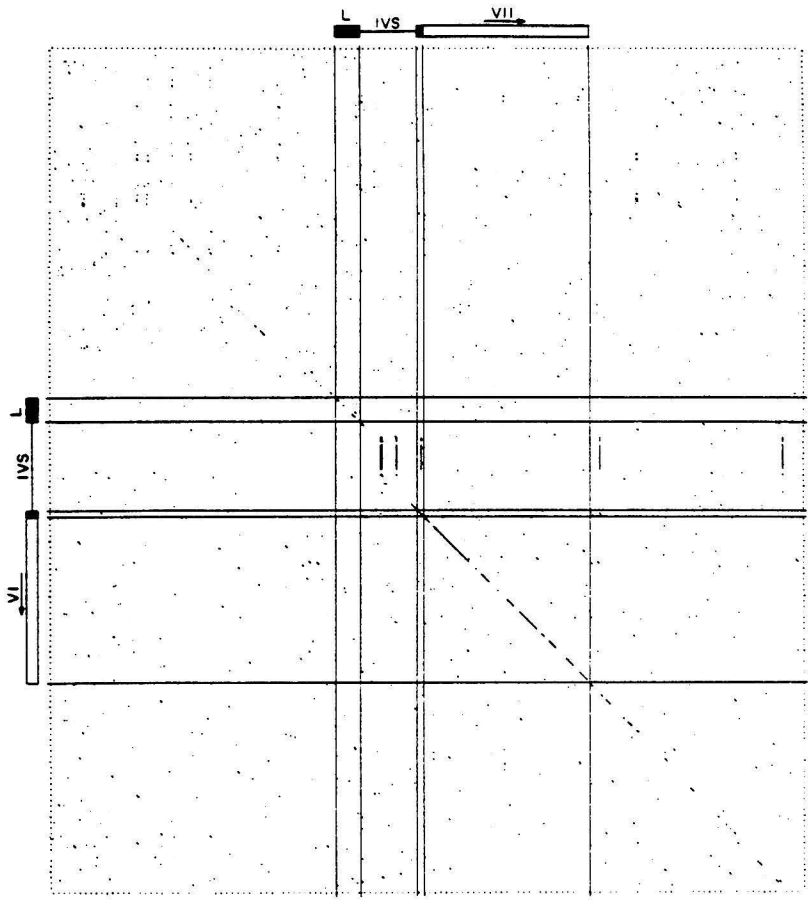
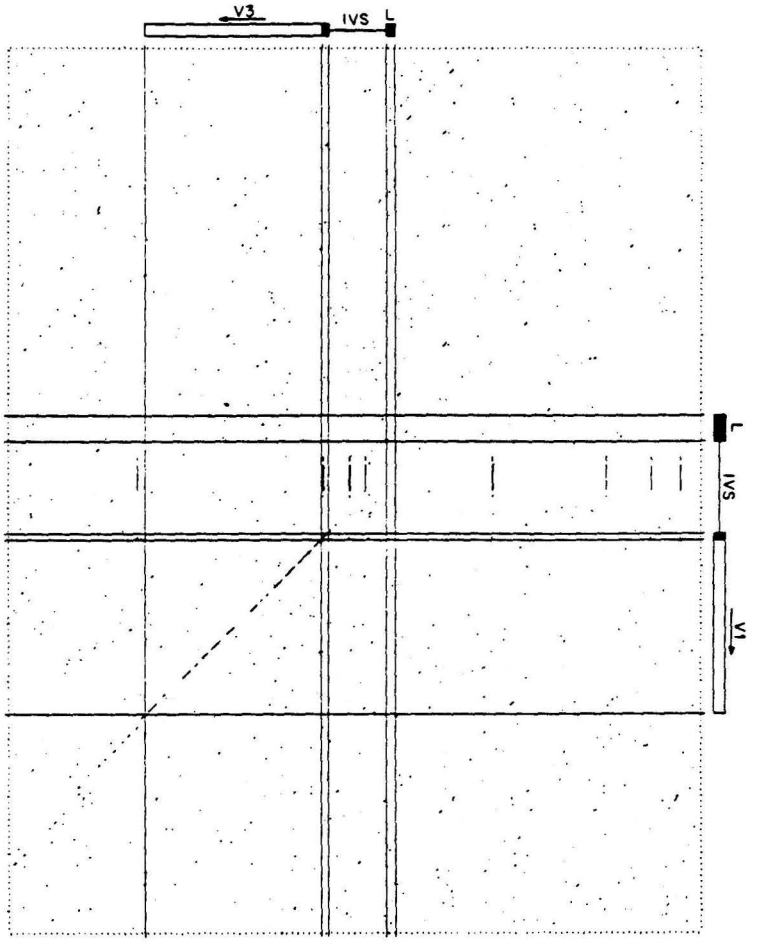
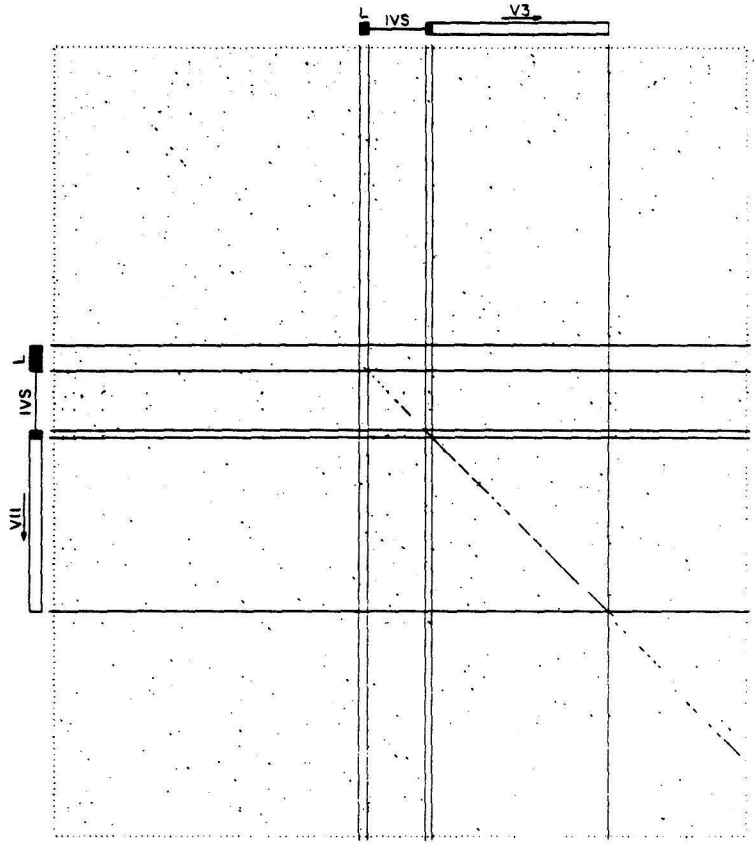

TATA
LEADER
LEADER
VARIABLE GENE SEGMENT
7
9

FIG. 3. Dot matrices comparing the members of the T15 family with each other. a) V11 vs. V13; b) V1 vs. V11; c) V3 vs. V1; d) V3 vs. V11; e) V3 vs. V13. All matrices are 5/5 comparisons (see text for details).


L
IVS
V13
L
IVS
VII


L
IVS
VII
L
IVS
VI


V3
IVS
L
L
IVS
V1


L
IVS
V3
L
IVS
VII

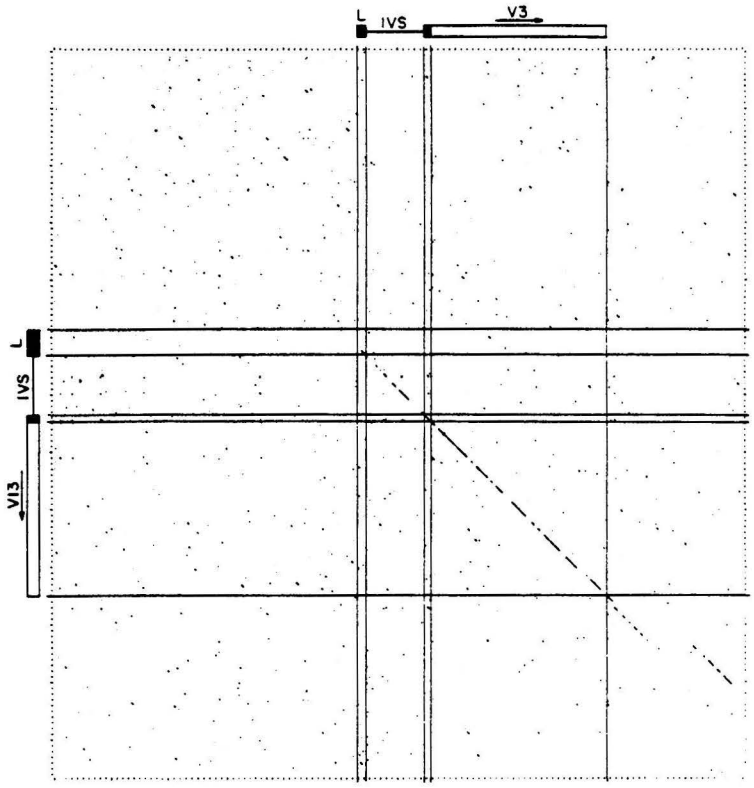

L
IVS
V3
L
IVS
VI3

FIG. 4. Mutation frequencies between the members of the T15 family subdivided into different regions. The 5' flank is defined as the region starting just 5' to the V_H box and continuing upstream. The 5' proximal region is defined as the region starting at the V_H box and ending at the cap site. The 3' flank is defined as the region starting just 3' to the V_H gene segment and continuing downstream. The mutation frequencies were calculated using the 3ST technique of Kimura (1981). Deletion and insertion events were not included in the calculations. Mutation frequency analysis within each comparison will not be affected, as the number of deletion or insertion events in the 5' and 3' regions are approximately equal. The comparisons with V1 in the intervening sequence, however, must be analyzed with caution. Error limits represent one standard deviation.

	5' FLANK	5' PROXIMAL	5'UT	L	IVS	V_H	3' FLANK
V1 VS V3	–	–	–	0.31 ± 0.08	0.32 ± 0.08	$K_1 = 0.084 \pm 0.030$ $K_2 = 0.154 \pm 0.042$ $K_3 = 0.131 \pm 0.038$ $K_A = 0.116 \pm 0.024$ $K_B = 0.147 \pm 0.084$	0.54 ± 0.07
V1 VS V11	1.22 ± 0.13	0.33 ± 0.09	0.33 ± 0.09	0.12 ± 0.05	0.30 ± 0.08	$K_1 = 0.082 \pm 0.026$ $K_2 = 0.130 \pm 0.039$ $K_3 = 0.119 \pm 0.037$ $K_A = 0.086 \pm 0.020$ $K_B = 0.164 \pm 0.055$	0.43 ± 0.05
V1 VS V13	1.04 ± 0.11	0.35 ± 0.09	0.36 ± 0.09	0.16 ± 0.06	0.28 ± 0.08	$K_1 = 0.084 \pm 0.030$ $K_2 = 0.191 \pm 0.048$ $K_3 = 0.156 \pm 0.044$ $K_A = 0.121 \pm 0.024$ $K_B = 0.226 \pm 0.068$	0.44 ± 0.06
V3 VS V11	–	–	–	0.22 ± 0.07	0.17 ± 0.04	$K_1 = 0.020 \pm 0.014$ $K_2 = 0.083 \pm 0.026$ $K_3 = 0.041 \pm 0.021$ $K_A = 0.044 \pm 0.014$ $K_B = 0.030 \pm 0.020$	0.15 ± 0.03
V3 VS V13	–	–	–	0.28 ± 0.09	0.18 ± 0.04	$K_1 = 0.041 \pm 0.021$ $K_2 = 0.097 \pm 0.033$ $K_3 = 0.085 \pm 0.031$ $K_A = 0.063 \pm 0.017$ $K_B = 0.110 \pm 0.043$	0.12 ± 0.03
V11 VS V13	0.20 ± 0.03	0.11 ± 0.04	0.09 ± 0.04	0.05 ± 0.03	0.05 ± 0.02	$K_1 = 0.030 \pm 0.018$ $K_2 = 0.073 \pm 0.027$ $K_3 = 0.085 \pm 0.030$ $K_A = 0.048 \pm 0.014$ $K_B = 0.111 \pm 0.044$	0.09 ± 0.02

FIG. 5. 3' flanking region of different germline V_H gene segments. The 7 bp and 9 bp rearrangement recognition signals are boxed and indicated. Two additional conserved sequences located in the 23 bp spacer sequence are overlined. The references for the sequences are: V11, V13, V3 and V1 (this paper); VH441 (Ollo *et al.*, 1981); V14B (S. Crews, unpublished); VH104, VH108A, VH108B, and VH111 (Givol *et al.*, 1981); VH105 (Cohen *et al.*, 1982); VH141 (Sakano *et al.*, 1980); VHIII (Rabbits *et al.*, 1980); V23 (Bothwell *et al.*, 1981).

```
                7                                  9
V11      1   ACACAGTGAGGGTACTTCAGTGTGAGCCTAGACACAAACCTCC TTGCAAAGGT GCTCAGGA CCAACAGGGGGCGCACA  TGACCAGAGAAAAAA  93
V13      1  GT.......................................... ......GG... ........ ..G.........A....  ...              83
V3       1  AC........ .....................T..A........... AC....GG.C. ........ ..............T...  .....GC.......G  87
V1       1  GC........ ..A.G..G...T.......C............. .A.....GG... .T..T... ...... .A........  ....T.TG......G  95
VH441    1   CC......... ..AA.TC....T..TA..C...TG...... AC.. TG....C.....CA. ...C.........ATG.  AA......C.C.G..  94
V14B     1   GG......... .AAGC...T.T.A...A....TC.......T GC.. C.G.T.TT..G...C  ...TT......A.....G  ..G.T.AGT...CC.  95
VH104    1   .......TT.TA..CA..TCC....TG.GT..G.....CT        G.A.A   G.....AG.TG..CT... A.TG.GA...  ....A.G.TT.  84
VH105    1   .......TT.TA..CA..TCC....TG.GT..G.....CT        G...A   G.....AG.TG..CT... A.TG.GA...  ....A.G.TT.  84
VH108A   1   .......TTACA.ACA..TCC....TG.GT..G.....CT        G....   G...C.AG.TT.CTT... A.TG.CAA..  GTT...G..T.  84
VH108B   1   .....C.TT.TA..CA...GC....TG.GT..G...A..        G.....   G...C.AGGTG.CTT.A. A.TG.CAA..  .TT...G..GT  84
VH111    1   .....C.TT.CA..CA..TCC....TG.GT..G..C.TCT        G...A   G.....AG.TG..CT.C. A.GG.GA...  ....A.G.TT.  84
VH141    1  AG.........A.G.C..A.........G.......TAC.TC.CTGC.G..AT.A...CA.  ...G...........TG.  G....CA..G.CTT  94
VH-III   1   GG.........AG G............C............ .....GG.CA  ...T...G...CA........T..                    76
V23      1   .......TT.CA..CA..TCC....AG.GT..G...A.CT          ..AT.AGT..C.AA.TG.CCT.A. A.TG.GGA..  .TC...G..GG  85
                    |         |         |         |         |         |         |         |         |         |

V11     94  AGT GCATTGCCTCT CCTTGTATTACTTAACATCTGAAAA TTCTTCTG     140
V3      88  ... ......T.... .T.C...C.TA.. ......G...TA             126
V1      96  .C. .T..GA..... T.A.....C....TTT.....G...G.......      142
VH441   95  CACA.G..C...C.                A.G...TGAGAAACA          123
V14B    96  G.AACAGGA..TA..GTT.ATAGAG.GA.C.GGGGA.GC..AAC.ACT.T     145
VH104   85  .TCTTT.G.CTTG..CAGAAA.TG..A..TTG.ATGTCC.TT.AT.         130
VH105   85  .TCCTT.GACT....CAGAAA.TG..A..CTG.GTGTCC.TT.AT.         130
VH108A  85  GTCGCTTGCAGACT.G.T.A.ATGC.G.C.TTTGAATGGTTG..T.GTG      133
VH108B  85  GTAGCTTGCAGG.T.G.T.A.ATGCTG...TTTGGGT..TTG..T..TG      133
VH111   85  .TCACTCAGAAT.GCT.AGAAAT.GTAA.TT.GAA..TCC.TG.A..        131
V23     86  TT.GCTTG.                                               94
                    |         |         |         |         |
```

FIG. 6. Comparison of the translated protein sequences of different germline V_H gene segments. The hypervariable regions according to Wu and Kabat (1970) are indicated. The references for the sequences are: V1 (this paper); V14B (S. Crews, unpublished); VH101 (Kataoka *et al.*, 1982); VH186-1 (Bothwell *et al.*, 1981; V108B (Givol *et al.*, 1981); Caiman VH (Litman *et al.*, 1983); VH26 (Matthyssens and Rabbitts, 1980); VH-III (Rabbitts *et al.*, 19080); VH-I (Rechari *et al.*, 1983); VH441 (Ollo *et al.*, 1981); VHA1 (D. Livant, unpublished).

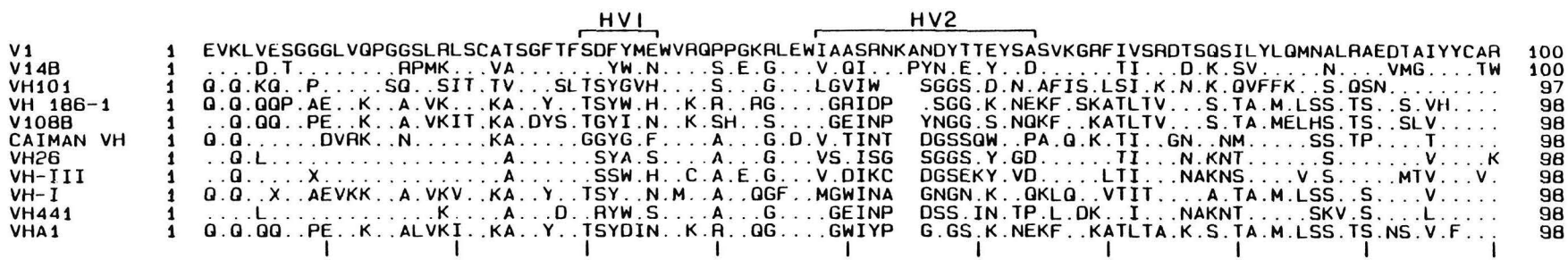

HV1
HV2
V1 1 EVKLVESGGGLVQPGGSLRLSCATSGFTFSDFYMEWVRQPPGKRLEWIAASRNKANDYTTEYSASVKGRFIVSRDTSQSILYLQMNALRAEDTAIYYCAR 100
V14B 1 ....D.T.....RPMK...VA.....YW.N....S.E.G...V.QI...PYN.E.Y..D.....TI...D.K.SV.....N....VMG....TW 100
VH101 1 Q.Q.KQ..P....SQ..SIT.TV...SLTSYGVH....S...G...LGVIW SGGS.D.N.AFIS.LSI.K.N.K.QVFFK..S.QSN..... 97
VH 186-1 1 Q.Q.QQP.AE..K..A.VK...KA..Y..TSYW.H..K.R..RG....GRIDP .SGG.K.NEKF.SKATLTV...S.TA.M.LSS.TS..S.VH.... 98
V108B 1 ..Q.QQ..PE..K..A.VKIT.KA.DYS.TGYI.N..K.SH..S....GEINP YNGG.S.NQKF..KATLTV...S.TA.MELHS.TS..SLV..... 98
CAIMAN VH 1 Q.Q.....DVRK..N.....KA.....GGYG.F....A...G.D.V.TINT DGSSQW..PA.Q.K.TI..GN..NM.....SS.TP....T..... 98
VH26 1 ..Q.L.....A.....SYA.S....A...G...VS.ISG SGGS.Y.GD.....TI...N.KNT.....S.....V....K 98
VH-III 1 ..Q.....X.....A.....SSW.H..C.A.E.G...V.DIKC DGSEKY.VD.....LTI...NAKNS....V.S.....MTV...V. 98
VH-I 1 Q.Q..X..AEVKK..A.VKV..KA..Y..TSY..N.M..A..QGF..MGWINA GNGN.K..QKLQ..VTIT....A.TA.M.LSS..S....V..... 98
VH441 1 ....L.....K....A...D..RYW.S....A...G....GEINP DSS.IN.TP.L.DK..I...NAKNT.....SKV.S....L..... 98
VHA1 1 Q.Q.QQ..PE..K..ALVKI..KA..Y..TSYDIN..K.R..QG....GWIYP G.GS.K.NEKF..KATLTA.K.S.TA.M.LSS.TS.NS.V.F... 98

FIG. 7. Wu-Kabat analysis for variability. Variability at a given amino acid position N is calculated as

$$\text{Variability}_N = \frac{\text{number of different amino acids at position N}}{\text{frequency of most commonly-occurring amino acid at position N}}$$

Hypervariable regions as defined by Wu and Kabat (1970) are indicated. Sequences used in this comparison are listed in Figure 6.

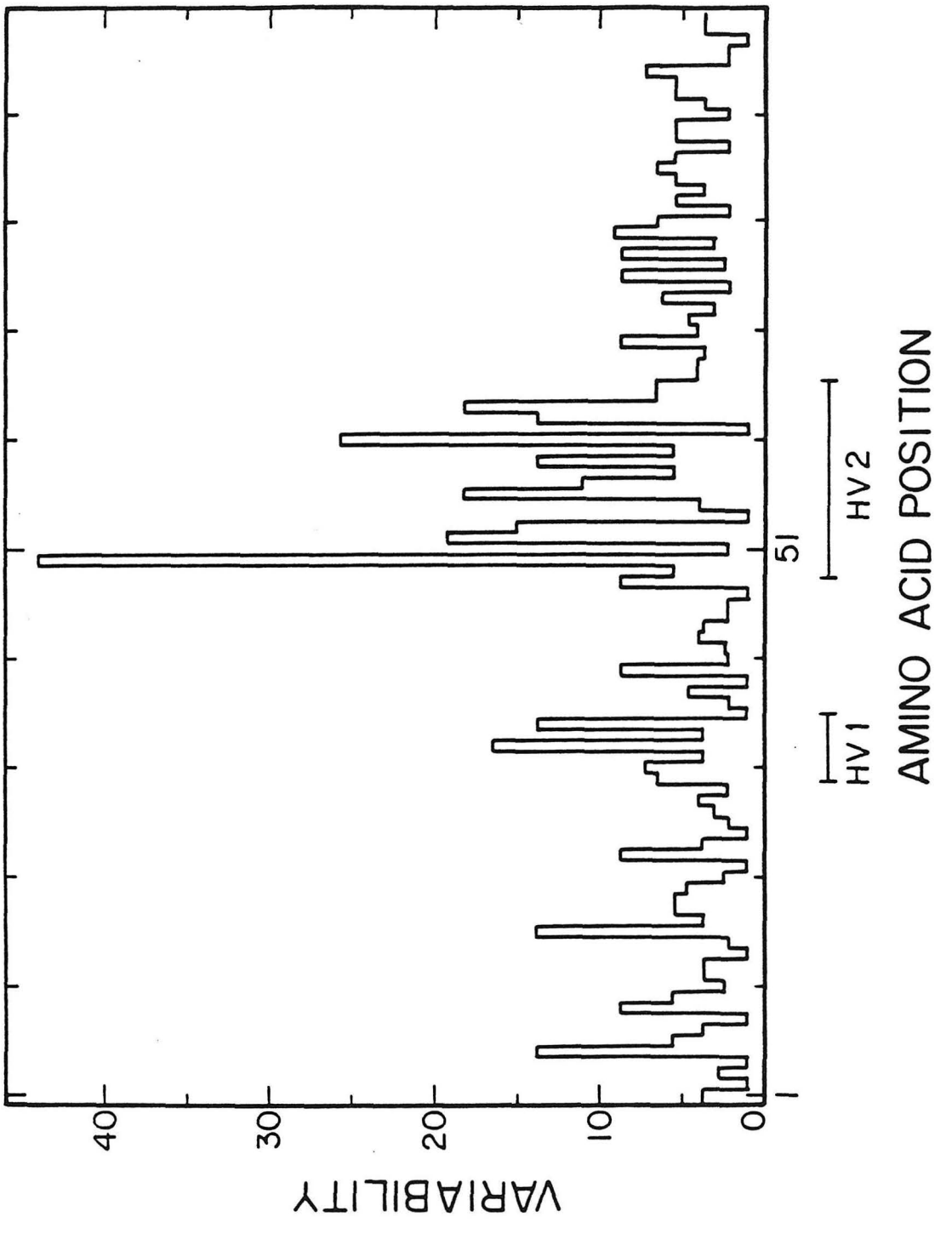

VARIABILITY
0
10
20
30
40
1
51
HV1
HV2
AMINO ACID POSITION

FIG. 8. 5' untranslated and flanking regions of different V_H gene segments. Gaps are inserted to maximize homologies. The cap site for V1 has been identified (the boxed "A"). The TATA and V_H boxes are boxed. References are: V11, V13, V1 (this paper); V108A, V108B, V104 and V111 (Givol *et al.*, 1981); V105 (Cohen *et al.*, 1982); BCL1 (Knapp *et al.*, 1982); Mu + 2 (Early *et al.*, 1982); V101 (Kataoka *et al.*, 1982); HA2 and HG3 (Rechavi *et al.*, 1983).


V11
V13
V1
V108A
V108B
V111
V104
V105
MU+2
HG3
HA2
" V_H "
" TATA "

FIG. 9. Conserved sequences in the 5' flanking regions of V_K regions and histone H2b genes. References are: T1 (Altenburger *et al.*, 1980), M173B (Max *et al.*, 1980), K2 (Nishioka and Leder, 1980), HK101 and HK102 (Bentley and Rabbitts, 1980), MOPC41 (Seidman *et al.*, 1979), M167 (Selsing and Storb, 1981), MPC 11 (Kelley *et al.*, 1982), MPC11 non-productive rearrangement (Seidman and Leder, 1980), histone H2B alignments (Harvey *et al.*, 1982). The conserved sequence was also found in the MOPC215 λ light chain (Wu *et al.*, 1982). The conserved core sequence is present in MPC11, but it is located 8 bp 3' to the TATA homology.

T1	GCAAT	13bp	TGATTTGCATGC	30bp	CATATA
M173B	CCAAT	17bp	TGATTTGCATAT	30bp	CATATA
K2			TGATTTGCATGT	37bp	CATAGT
HK101	CCCAT	8bp	TCATTTGCATGT	37bp	AATAGG
HK102	CCCAT	8bp	TCATTTGCATGT	37bp	AATAGG
MOPC41	CCAAT	17bp	TAATTTGCATAC	30bp	TATATAA
M167	?	?	G TGATTTGCATAC	40bp	CAAAAG
MPC11	TCCAT		56bp		TATAA
MPC11 (NON-PROD)			TGATTTGCATA	0bp	AAATAA
CH02	ATCCAATCA	22bp	C CTCATTTGCATAC	6bp	TATAAATA
CH01/ CH05	GACCAATCA	23bp	CTCATTTGCATAG	6bp	TATAAATA
X. LAEVIS	TTACAAGAT	29bp	CTTATTTGCATGG	6bp	TATAAAA
S. PURPURATUS	GACCAATGA	17bp	CTCATTTGCATAC	29bp	TATAAAAA
P. MILIARIS	GCCCAATGA	18bp	CTCATTTGCATAC	31bp	TATAAAAA

CHAPTER THREE

THE HUMAN T-CELL ANTIGEN RECEPTOR IS ENCODED BY VARIABLE, DIVERSITY, AND JOINING GENE SEGMENTS THAT REARRANGE TO GENERATE A COMPLETE V GENE

Published in *Cell*

Cell, Vol. 37, 393-401, June 1984, Copyright © 1984 by MIT 0092-8674/84/060393-09 $02.00/0

The Human T Cell Antigen Receptor Is Encoded by Variable, Diversity, and Joining Gene Segments That Rearrange to Generate a Complete V Gene

Gerald Siu,* Stephen P. Clark,† Yasunobu Yoshikai,† Marie Malissen,* Yusuke Yanagi,† Erich Strauss,* Tak W. Mak,† and Leroy Hood*
*Division of Biology
California Institute of Technology
Pasadena, California 91125
†Department of Medical Biophysics
University of Toronto
Ontario Cancer Institute
Toronto, Canada M4X 1K9

Summary

A cDNA clone YT35, synthesized from poly(A)* RNA of the human T cell tumor Molt 3, exhibits homology to the variable (V), joining (J), and constant (C) regions of immunoglobulin genes. We have isolated and sequenced the germ-line V and J gene segment counterparts to YT35 from a human cosmid library, and these failed to encode 14 nucleotides of the cDNA clone between the V and J regions. We postulate that these 14 nucleotides are encoded by a third gene segment analogous to the diversity (D) gene segments of immunoglobulin heavy chain genes. This T cell antigen receptor V gene appears to be assembled from three gene segments, V, D, and J, and accordingly most closely resembles immunoglobulin heavy chain V genes.

Introduction

Characterization of the T cell antigen receptor has been an elusive goal of molecular immunologists for many years (e.g. Kraig et al., 1983; Kronenberg et al., 1983a, 1983b). The first real progress in this regard came within the last year when several groups produced clonally specific antibodies that reacted with cell surface T cell antigen receptors (Acuto et al., 1983; Kappler et al., 1983; McIntyre and Allison, 1983; Samelson et al., 1983). These molecules are heterodimers, composed of two disulfide-linked polypeptide chains, α and β. The α and β chains range in molecular weight from 42 to 45 kilodaltons in mice and from 39 to 49 kilodaltons in man. Comparative peptide maps of several α chains and several β chains demonstrate that they have constant and variable peptides, presumably encoded by distinct V and C regions (Kappler et al., 1983; McIntyre and Allison, 1983). Since the constant peptides of the α and β chains are different, they are presumably encoded by distinct α and β gene families.

Recently, two groups have used differential or subtractive screening methods to isolate cDNA clones that appear to encode one of the chains of the T cell antigen receptor (Hedrick et al., 1984a, 1984b; Yanagi et al., 1984). A comparative sequence analysis of the human and several mouse cDNA clones demonstrates that they contain sequences similar to immunoglobulin V, J, and C gene segments and their gene products have the same centrally positioned disulfide bridges seen in the V and C regions of immunoglobulins. The human and mouse T cell receptor cDNAs are 82% homologous in their C regions. This degree of similarity among immunoglobulins from two different species would suggest that both are members of the same gene family, and we presume the same is true of the human and mouse T cell receptor genes.

Two preliminary observations suggest that the β gene family of the T cell receptor is detected by the mouse and human cDNA probes. First, recent studies by Reinherz and his coworkers have demonstrated that the 12 N-terminal amino acids of a human T cell β chain are identical with those of the translated human cDNA clone YT35 (Acuto et al., 1984). Second, an analysis of the ^{125}I-labeled tryptic peptides of the α and β chains from mouse T cells suggests that only the β chain has a tryptic peptide distribution consistent with the translated sequences of the recently published mouse cDNA clone (Hedrick et al., 1984b; J. Kappler and P. Marrack, personal communication). We therefore believe that the human YT35 cDNA is derived from the β gene and have designated sequences homologous to this clone V_β and J_β respectively.

The genes encoding the T cell antigen receptors resemble immunoglobulin genes, whose structures and rearrangements are well understood (Honjo, 1983; Tonegawa, 1983). Immunoglobulins are coded for by three families of genes, λ, κ, and heavy chain, which are located on different chromosomes. The genes for light chains, λ and κ, have distinct variable (V) and joining (J) gene segments that rearrange into a contiguous DNA sequence to give complete V_L genes. Similarly, heavy chain genes have three gene segments, V, diversity (D), and J, which rearrange to generate V_H genes. These DNA rearrangement events appear to be mediated by recognition signals that lie immediately to the 3′ side of the V gene segments, the 5′ and 3′ sides of the D gene segments, and the 5′ side of the J gene segments. These recognition signals allow precise definition of the corresponding boundaries of these gene segments (Early et al., 1980).

We report the genomic organization of the coding elements of a human V_β gene. Using the cDNA clone YT35 to screen human genomic libraries, we have isolated and sequenced the germ-line V and J gene segments that most likely encode the V region of YT35. The presence of recognition sequences for DNA rearrangement permits precise definition of the 3′ end of the V gene segment and the 5′ end of the J gene segment. These gene segments fail to encode 14 nucleotides between the V and J regions of the YT35 cDNA. Since a second human cDNA clone, YTJ-2, which was isolated from the human T cell tumor Jurkat, contains an identical 14 nucleotides at the V–J boundary, it appears that this region is encoded by a separate gene segment similar to the D gene segments of immunoglobulin heavy chains. The germ-line V and J gene segments contain recognition sequences for DNA rearrangement similar to their immunoglobulin coun-

terparts. Thus the organization of the V_β gene of the T cell antigen receptor appears to resemble most closely that of the immunoglobulin heavy chain V gene.

Results and Discussion

Clones Containing β Gene Segments of the YT35 cDNA Clone

We have attempted to identify the germ-line V and J gene segments encoding the YT35 cDNA by using stringent hybridization conditions with the YT35 probe to identify the V gene segment and with a synthetic 30-mer oligonucleotide to identify the J gene segment. A Southern blot of germ-line DNA with the full length YT35 cDNA clone reveals four distinct bands with a Bam HI digest (Figure 1A). Separate Southern blots of human DNAs with V_β and C_β region probes subcloned from the YT35 cDNA clone demonstrate that these bands represent four distinct V_β genes (2.0, 3.3, 6.5, and 23 kb) and two C_β genes which identify a 23 kb Bam HI band that comigrates with a V_β-encoding restriction fragment (Figures 1B and 1C). The V_β gene segments of the YT35 gene family are denoted $V_{\beta M3\text{-}1}$ through $V_{\beta M3\text{-}4}$, beginning with the 2.0 kb fragment in the Bam HI digest (Figure 1). (M3 denotes Molt 3, the tumor from which the YT35 cDNA was originally isolated.) A cosmid library constructed from human sperm DNA was screened with the YT35 cDNA clone, and cosmid clones containing the $V_{\beta M3\text{-}1}$, $V_{\beta M3\text{-}2}$, and $V_{\beta M3\text{-}4}$ clones were isolated.

To approach the question of which germ-line $V_{\beta M3}$ gene segment corresponded to the V region of the YT35 cDNA clone, we used Southern blot analyses of human germ-line DNA with the subcloned V region of the YT35 cDNA under stringent hybridization conditions. Only V bands $V_{\beta M3\text{-}1}$ and $V_{\beta M3\text{-}2}$ remained intense, suggesting that they are more similar to the YT35 V region than are the $V_{\beta M3\text{-}3}$ and $V_{\beta M3\text{-}4}$ gene segments (data not shown). Restriction map analyses demonstrate that the $V_{\beta M3\text{-}2}$ gene segment shares several restriction sites with the YT35 V region that are lacking in the $V_{\beta M3\text{-}1}$ gene segment (data not shown). Therefore, we subcloned the $V_{\beta M3\text{-}2}$ gene segment and determined its sequence (Figure 2A).

A C_β-containing cosmid clone was isolated from a human osteosarcoma (U2 OS) DNA library. The C_β cosmid clone hybridized to a synthetic 30-mer DNA probe for the J region of the YT35 cDNA clone. The J gene segment homologous to the YT35 cDNA clone is approximately 3.8 kilobases from the C_β gene. The region of the cosmid clone hybridizing to the synthetic J probe was subcloned as a Sma I–Pvu II fragment and its sequence determined (Figure 2B).

A portion of the variable region sequence encoded by the YTJ-2 cDNA clone from the human T cell tumor Jurkat was also determined (Figure 2C).

The Germ-Line V Gene Segment Has Leader and Variable Exons

The nucleotide and translated protein sequences of the germ-line V gene segment are compared to those of the complete V sequence of cDNA clone YT35 and to the partial sequence of cDNA clone YTJ-2 (Figure 3). The V gene segment contains two exons of 49 and 295 base pairs, separated by an intron of 100 base pairs.

The smaller 5′ exon encodes a hydrophobic and uncharged sequence that is the presumed leader or signal peptide. The larger 3′ exon encodes most of the variable region. The processed human β chain has an N-terminal glycine residue (Acuto et al., 1984), hence the variable exon encodes a V segment of 94 codons. The first half-cysteine residue of the highly conserved disulfide bridge occurs at position 21—as contrasted with position 22 in V_H regions and position 23 in V_L regions. The disulfide bridge is comparable in size to those of immunoglobulins. The β gene exons and intron are very similar in size to the immunoglobulin V gene segments, although the sequence homologies are quite low except for highly conserved

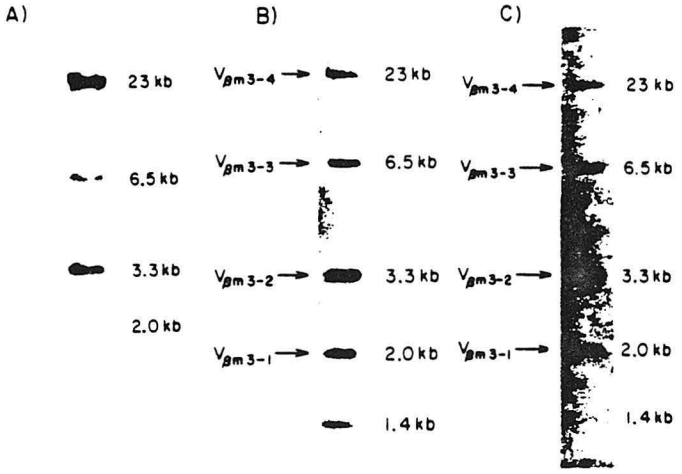


Figure 1. A Bam HI Southern Blot of Human DNA from the B Cell Clone FQ Hybridized against the YT35 cDNA

(A) A Southern blot with the complete cDNA. The sizes of the bands are indicated. The 2.0, 3.3, and 6.5 kb bands correspond to $V_{\beta M3\text{-}1}$, $V_{\beta M3\text{-}2}$, $V_{\beta M3\text{-}3}$, respectively. The 23 kb band contains both J–C clusters on one fragment and $V_{\beta M3\text{-}4}$ on another fragment that comigrates with the J–C cluster fragment. (B) Southern blot with the subcloned V region probe under normal hybridization conditions (three complete washes at 68°C with 3× SSC and three washes with 1× SSC). Occasionally a 1.4 kb band was seen. (C) Southern blot with the subcloned V region probe under stringent hybridization conditions (three complete washes at 68°C with 3× SSC, three washes with 1× SSC, and three washes with 0.1× SSC).

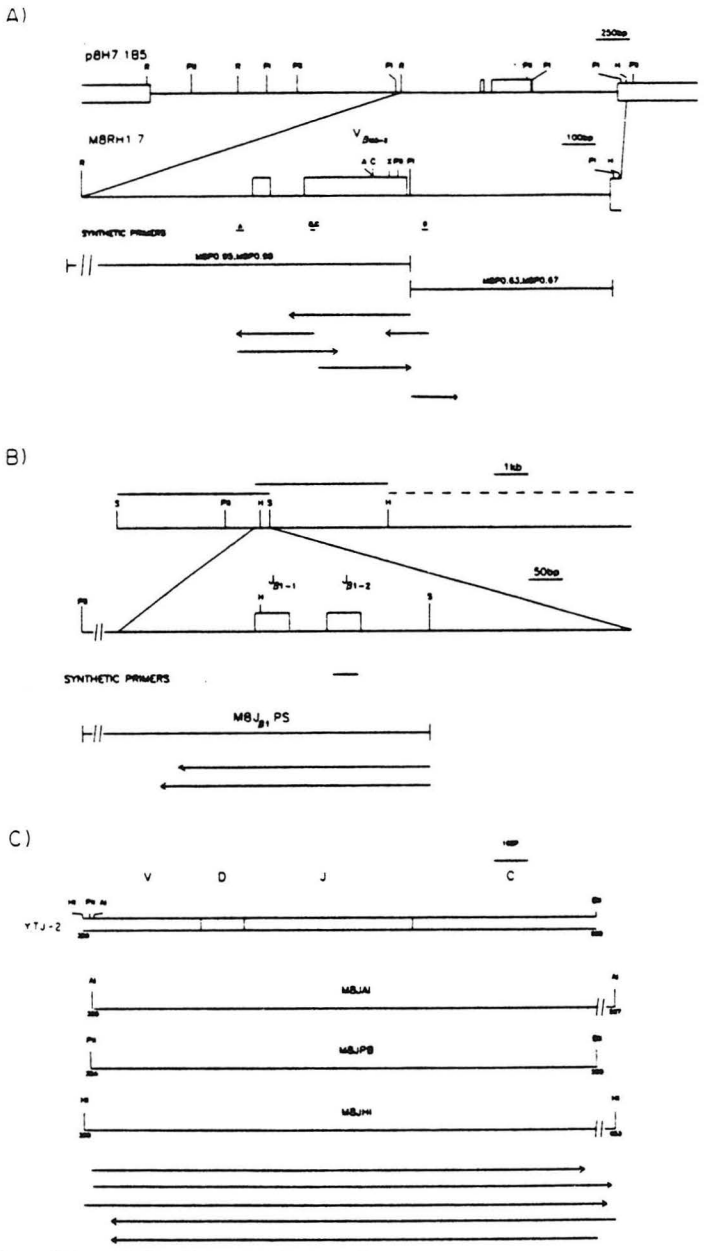


Figure 2. V_β and J_β Gene Clones and Sequencing Strategies

(A) A restriction map of clone p8H7.1B5, the pUC8 subclone of the $V_{\beta MG-2}$ gene segment. M8RH1.7, M8P0.95, M8P0.98, M8P0.63, and M8P0.67 are M13mp8 subclones of the $V_{\beta MG-2}$ gene segment and the 3′-flanking region. The boxed region at either end of the p8H7.1B5 map is the pUC8 vector. The abbreviations for the enzymes are as follows: A, Ava I; C, Cla I; H, Hind III; PI, Pst I; PII, Pvu II; R, RI; X, Xho II. Synthetic primers utilized for sequencing are indicated by horizontal lines. Arrows indicate the sequencing strategy.

(B) A restriction map of the $J_{\beta 1}$ PS clone in an M13mp9 subclone of the 1.0 kb Pvu II–Sma I fragment containing the two $J_{\beta 1}$ fragments. The abbreviations for the enzymes are as follows: S, Sma I; H, Hind III; PII, Pvu II; R, Eco RI. Arrows indicate sequencing strategy. The solid lines above the map indicate regions of hybridization with the J-region-specific oligonucleotide. The dashed line indicates hybridization with the C-region-specific probe from the YT35 cDNA.

(C) Partial restriction map of the Jurkat cDNA clone YTJ-2. M8JAI, M8JPB, and M8JHI are M13mp8 subclones of the V–D–J region of the β chain expressed by the Jurkat tumor cell line. The abbreviations for the enzymes are as follows: AI, Alu I; BII, Bgl II; HI, Hinf I; PII, Pvu II. Arrows indicate sequencing stragegy. As the Jurkat clone is virtually identical with the YT35 cDNA clone, the numbering indicated corresponds to the numbering of the YT35 cDNA in Figure 3 of Yanagi et al. (1984).

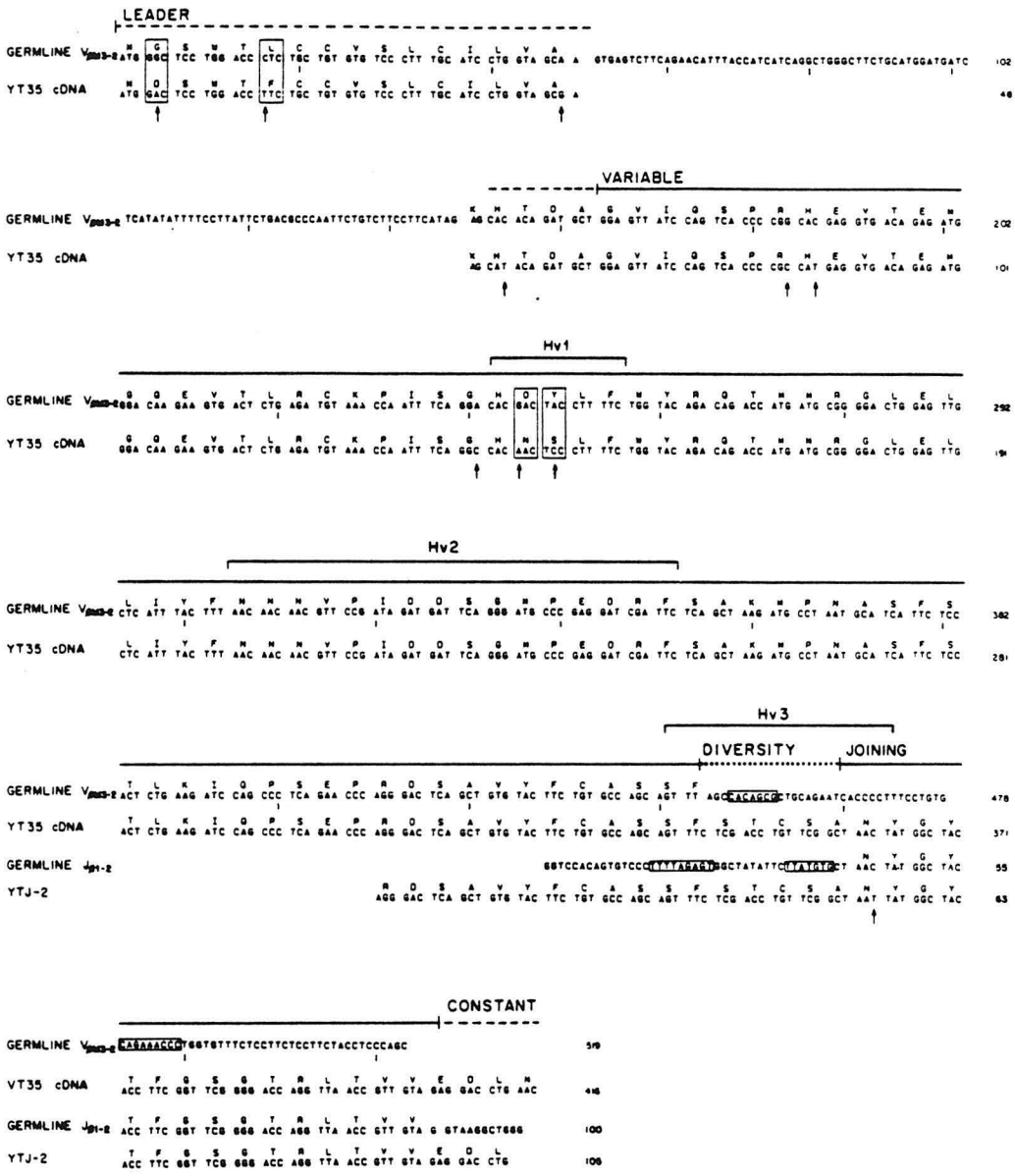


Figure 3. DNA Sequences of Germ-Line V_β and J_β Gene Segments and Rearranged V_β Genes

The leader, variable, diversity, joining, and constant gene regions are indicated. Positions that are analogous to the three hypervariable regions of V_H genes are indicated as Hv1, Hv2, and Hv3. Arrows indicate the sites of nucleotide substitutions between germ-line and rearranged sequences. Boxes indicate codons with replacement changes.

residues at the protein level (e.g., half-cysteines at codons 21 and 90 and a conserved sequence around the second half-cysteine).

The $V_{\beta M3\text{-}2}$ gene segment differs by nine nucleotide substitutions from the YT35 cDNA (97% homology), and its 3′ sequence is identical with that of the YTJ-2 cDNA clone. Three of these substitutions are in the leader exon and six are in the variable exon (Figure 3).

Approximate RNA splicing signals delineate the 3′ end of the leader exon (GT) and the 5′ end of the V exon (AG) (Breathnach and Chambon, 1981). As with the genes in the immunoglobulin and major histocompatibility complex

gene families, RNA splicing always occurs between the first and second bases of the junctional codon. This RNA splicing pattern is another suggestion, in addition to sequence homology, that the class I, class II, immunoglobulin, and T cell receptor gene families are evolutionarily related.

The general organization of the V_β gene segment of the T cell receptor is clearly similar to the organization of immunoglobulin V gene segments.

One J_β Gene Segment Encodes the J Region of the YT35 cDNA Clone

The nucleotide sequence of a portion of the Sma I–Pvu II fragment isolated from a human C_β clone and hybridizing to a synthetic oligonucleotide for the J_β region of YT35 is presented in Figure 4A. Dot matrix analysis of this sequence against the J_β sequence of the YT35 cDNA clone identified two homologous regions, which are denoted $J_{\beta1\text{-}1}$ and $J_{\beta1\text{-}2}$. (There are two human C_β genes denoted $C_{\beta1}$ and $C_{\beta2}$. Each probably has its own cluster of J gene segments designated $J_{\beta1}$ and $J_{\beta2}$. The J gene segments analyzed here probably come from the $J_{\beta1}$ gene cluster; M. Malissen, S. Clark, and Y. Yanagi, unpublished results.) The $J_{\beta1}$ gene segments are 48 base pairs long and code for 15 amino acids. They differ by 5 of 15 amino acid residues (67% homology) and by 21 of 48 nucleotides (56% homology). It is striking that 17 of the 21 nucleotide substitutions and all 5 of the amino acid replacements occur in the 5′ half of the J_β region (Figure 4B). Additional analysis of J_β gene segments will be necessary to determine whether this is a consistent pattern with possible functional implications.

The J_λ (13 codons; Blomberg and Tonegawa, 1982), J_κ (13 codons; Max et al., 1981; Hieter et al., 1982), and J_H (15 or 17 codons; Early et al., 1980; Gough and Bernard, 1981) gene segments are similar in size to their $J_{\beta1}$ counterparts. The two J_β gene segments are slightly more homologous at the protein level to J_λ (62%–69% homology) than to J_κ (46%–61% homology), J_H (31%–46% homology), or themselves (59% homology). On the DNA level, the 3′ region of the $J_{\beta1}$ gene segments appears to be more homologous to J_λ (66%), J_κ (64%–71%), and J_H (59%) than the 5′ region is. The 5′ region bears a strong homology to the D regions of the immunoglobulin heavy chain. The 5′ region of $J_{\beta1\text{-}1}$ (nucleotides 64–76 of Figure 4A) shares 8 of 12 nucleotides with an immunoglobulin germ-line D segment from the FL16.1 family, while the 5′ region of $J_{\beta1\text{-}2}$ (nucleotides 200–214 of Figure 4A) shares 11 of 14 nucleotides with Sp2.6, a D segment from the Sp2 family (Kurosawa and Tonegawa, 1982). As mentioned above, this 5′ region also is quite different in the $J_{\beta1\text{-}1}$ and $J_{\beta1\text{-}2}$ gene segments. These observations may have interesting implications for the evolution of D and J gene segments in T and B cells.

The RNA splicing signals at the 3′ side of these J_β gene segments are the canonical splice-recognition donor signals (Breathnach and Chambon, 1981). The RNA splicing occurs between the first and second bases of the codon, as noted above for RNA splicing of the V gene segment.

Possible Somatic Mutation and Polymorphism in T Cell β Genes

A comparison of the nucleotide sequences of a germ-line $V_{\beta M3\text{-}2}$ gene segment and its rearranged cDNA counterpart YT35 (Figure 3) demonstrates nine nucleotide differences between these genes. Three of these differences occur in the first exon and six in the second exon. Two of the substitutions in the first exon and two in the second exon lead to amino acid replacements (Figure 3). In addition, the YTJ-2 cDNA appears to have one silent substitution in the J region relative to the germ-line J gene segment.

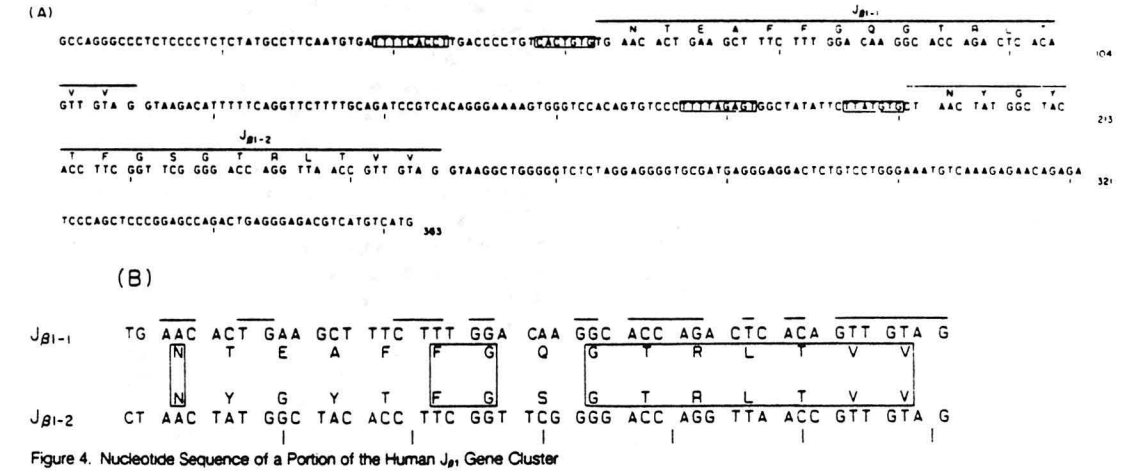


Figure 4. Nucleotide Sequence of a Portion of the Human $J_{\beta1}$ Gene Cluster

(A) The nucleotide sequence of a portion of the human $J_{\beta1}$ gene cluster. The coding regions are translated and the recognition sequences are boxed.
(B) Comparison of $J_{\beta1\text{-}1}$ and $J_{\beta1\text{-}2}$ sequences. The nucleotide identities are denoted with a solid line above the $J_{\beta1\text{-}1}$ sequence. Amino acid identities are boxed. The synthetic oligonucleotide for the YT35 J region extends from nucleotide 216 to 246.

These base substitutions could arise by several mechanisms. A V_β gene that is closely related to, but not the germ-line counterpart of, the YT35 V region may have been sequenced. The Southern blotting and restriction map data discussed above would argue against this possibility, as does a preliminary sequence analysis of the very similar $V_{\beta 3-1}$ gene segment (E. Strauss and G. Siu, unpublished observations). Or these differences could arise as a result of polymorphism in the human population. The YT35 and YTJ-2 cDNA clones as well as the germ-line $V_{\beta 3-2}$ gene were clearly isolated from the DNAs of different individuals. Several examples are known in immunoglobulins of alleles that have multiple nucleotide substitutions (Ollo and Rougeon, 1983). Finally, some or all of these differences might arise by the somatic hypermutation mechanism known to operate in immunoglobulin genes (Kim et al., 1981). In this regard, it is interesting that the two replacement substitutions in the YT35 cDNA V region occur in a region analogous to the first hypervariable region of immunoglobulin heavy chain genes. Perhaps these differences are selected by antigen (Kim et al., 1981; Crews et al., 1981) or by the need to mutate away from anti-self (Jerne, 1974).

The β Gene Family of T Cell Receptors Appears to Have D Gene Segments

A comparison of the germ-line V_β and J_β gene segments with the YT35 cDNA clone demonstrates that there are 14 nucleotides in the junctional region between the V and J gene segments that are not encoded by either germ-line gene segment (Figure 3). There are three possible explanations for the additional nucleotides seen in the cDNA clones. First, these additional nucleotides may represent a D gene segment analogous to those found in the immunoglobulin heavy chain gene family. We think this is the most likely explanation. Second, these nucleotides might arise from a somatic mutational mechanism similar to the N-terminal-region diversity postulated to be generated by random nucleotide addition to the junctional ends of the V, D, or J gene segments—an enzymatic addition presumably mediated by an enzyme such as terminal deoxynucleotidyl transferase (Baltimore, 1974; Alt and Baltimore, 1982). This second possibility is extremely unlikely because the nucleotide sequence of the cDNA clone derived from a separate T cell tumor is identical with that of YT35 at 14 of the 14 bases in this region (Figure 3). The Jurkat T cell tumor line was derived independently and differs from Molt 3 in HLA haplotype (J. Falls and T. Mak, unpublished observations) as well as in the nucleotide sequence of the J region (Figure 3). It is extremely unlikely that any type of somatic mutational mechanism could generate identical sequences of 14 consecutive nucleotides in two independently derived V_β genes. Third, a different germ-line V_β gene segment might have rearranged to form the YT35 cDNA variable region, and this gene segment might be 14 nucleotides longer than the $V_{\beta 3-2}$ gene segment. This explanation appears unlikely in that the ten immunoglobulin V gene families studied to date each have V gene segments of identical length. These include seven V_H families (Bothwell et al., 1981; Crews et al., 1981; Crews, unpublished data; Givol et al., 1981; Rechavi et al., 1982; Loh et al., 1983; Perlmutter et al., 1984; Perlmutter et al., unpublished data) and three V_κ families (Seidman et al., 1978; Bentley and Rabbitts, 1980, 1983). Accordingly, even if we have sequenced a V gene segment similar to but not identical with the germ-line YT35 gene segment, both are in the same V gene family and should be the same length. Thus it appears likely that the extra 14 nucleotides between the V and J regions of the YT35 cDNA represent an additional gene segment comprising part of the T cell receptor variable genes.

The DNA Recognition Sequences for the DNA Rearrangements of V_β and J_β Gene Segments Are Homologous to Their Immunoglobulin Counterparts but Differ in Their Disposition

Immunoglobulin genes have recognition sequences that presumably mediate DNA rearrangements. These sequences lie to the 3′ side of the V, to the 5′ and 3′ sides of the D, and to the 5′ side of the J gene segments (Figure 5). These recognition sequences are composed of three elements (Figure 5A). At one end is a highly conserved heptamer and at the other is a somewhat less conserved nanomer; a much less conserved spacer sequence of either 11 nucleotides (approximately one turn of the DNA helix) or 22 nucleotides (two turns of the DNA helix) separates the heptamer and nanomer elements. V-gene assembly always joins a one-turn recognition element to a two-turn element in the three immunoglobulin gene families (Figure 5B). Two observations suggest that these recognition elements play an important role in DNA rearrangements. First, they have been highly conserved over more

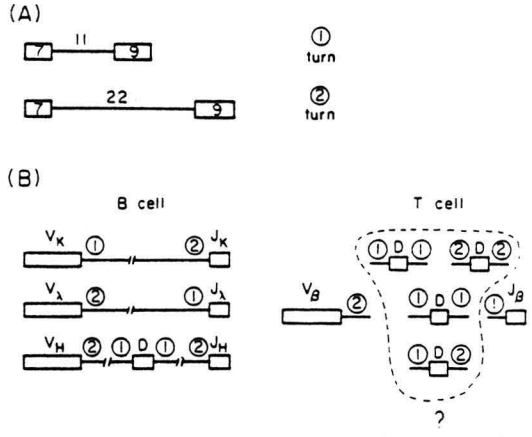


Figure 5. Schematic Diagram of the Recognition Sequences for DNA Rearrangements in B Cell and T Cell Receptor Genes

(A) A model of the conserved heptamer and nanomer sequences with 11 and 22 nucleotide spacer sequences.

(B) A diagrammatic representation of the recognition signals for immunoglobulin κ, λ, and heavy chain genes and the β T cell receptor genes.

than 350 million years of vertebrate evolution (Litman et al., 1983), suggesting that they are highly selected for an important function. Second, they are strategically placed at precisely the points of DNA rearrangement.

The β gene family of T cell receptors has elements that are very similar to the immunoglobulin recognition sequences (Figure 6). The $V_{\beta M3\text{-}2}$ gene segment has a two-turn recognition sequence including heptamer and nanomer elements identical with their counterparts among the immunoglobulin gene families. The two recognition sequences for J_β gene segments are both one-turn elements. The J_β heptamer is highly conserved in that it is identical with the mouse $J_{\lambda 3}$ heptamer sequence. Thus the V_β heptamer and nanomer as well as the $J_{\beta 1}$ heptamers are virtually indistinguishable from those found adjacent to immunoglobulin genes. In contrast, the $J_{\beta 1}$ nanomer does not have an exact immunoglobulin counterpart. These observations suggest that the recognition mechanisms and enzymatic machinery for immunoglobulin and T cell receptor gene rearrangements are very similar. This contention is supported by the fact that T cells have occasional D_H–J_H immunoglobulin gene segment rearrangements (Kurosawa et al., 1981). This observation implies that the converse may also be found, D_β–J_β rearrangements in B cells. Obviously the DNA rearrangements in T and B cells must exhibit appropriate tissue specificity.

The β gene family with V, D, and J gene segments appears to be more similar to the immunoglobulin heavy chain gene family than to the light chain families. However, the β gene family differs strikingly in the disposition of recognition sequences from the immunoglobulin heavy chain genes in that the J_β gene segment has a one-turn rather than a two-turn recognition element (Figure 5b). There are four possible explanations for this interesting difference in the disposition of recognition sequences. One explanation is that the β gene family does not have D gene segments, and as with the immunoglobulin λ gene family, a V two-turn recognition sequence joins with a J one-turn recognition sequence (Figure 5a). This explanation would require that identical sequences between the V and J gene segments in the YT35 and YTJ-2 cDNA clones arise by somatic mutation, a seemingly unlikely possibility. The second explanation is that the D gene segments are of two types, those with one-turn recognition signals on either side and those with two-turn recognition sequences on either side. Variable gene formation would always require D–D joining events if DNA rearrangements are mediated by one-turn to two-turn joining events as in B cells. Third, the D_β recognition elements may be identical with their immunoglobulin heavy chain counterparts (Figure 5b). This organizational disposition would require a new recognition system for joining like recognition elements (e.g., one-turn D_β to one-turn J_β). One attractive feature of this model is that it provides a mechanism by which immunoglobulin and T cell receptor gene rearrangements could be differentially controlled during development. That is, if T cells need a special recognition or joining systems for one-turn to one-turn joining, then these systems may only be turned on in T cells. Finally, the D_β gene segment could have 5′ one-turn and 3′ two-turn recognition elements. In this model, the basic rules for joining and recognition are precisely the same as those for immunoglobulin genes. In both the second and fourth models, if the β gene family has D gene segments and employs the one-turn to two-turn joining rules, then additional diversity can be generated in the β T cell receptor genes by at least two distinct mechanisms. First, D gene segments may join to one another. Second, V gene segments may join directly to J gene segments because they have two-turn and one-turn recognition elements respectively. If either of these models is correct, the T cell receptor genes may have considerably more potential for diversification through combinatorial DNA rearrangements than their immunoglobulin counterparts have. Resolution of these competing hypotheses will require the isolation and structural characterization of germline D_β gene segments and their putative recognition elements.

The V_β Gene Is Similar to Immunoglobulin V_H Genes

The V_β gene of the T cell antigen receptor appears to resemble most closely the immunoglobulin V_H gene in that it has three gene segments, V_β, D_β, and J_β, which rearrange

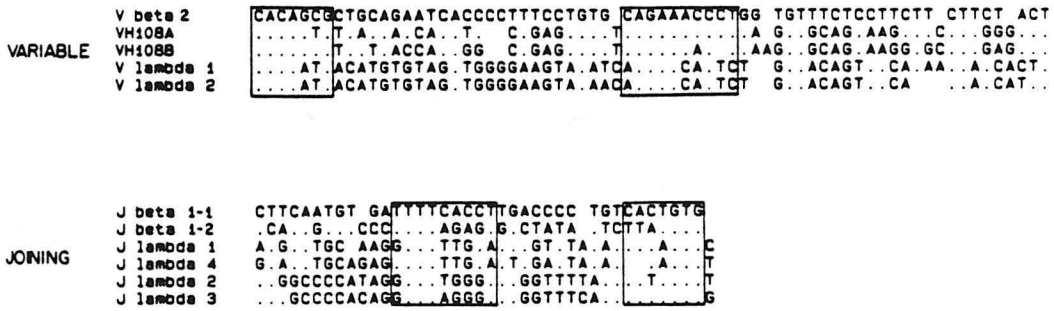


Figure 6. A Comparison of the Recognition Sequences of T Cell Gene Segments with Their B Cell Counterparts

The recognition sequences are boxed. V_H and V_λ denote immunoglobin heavy and λ recognition sequences.

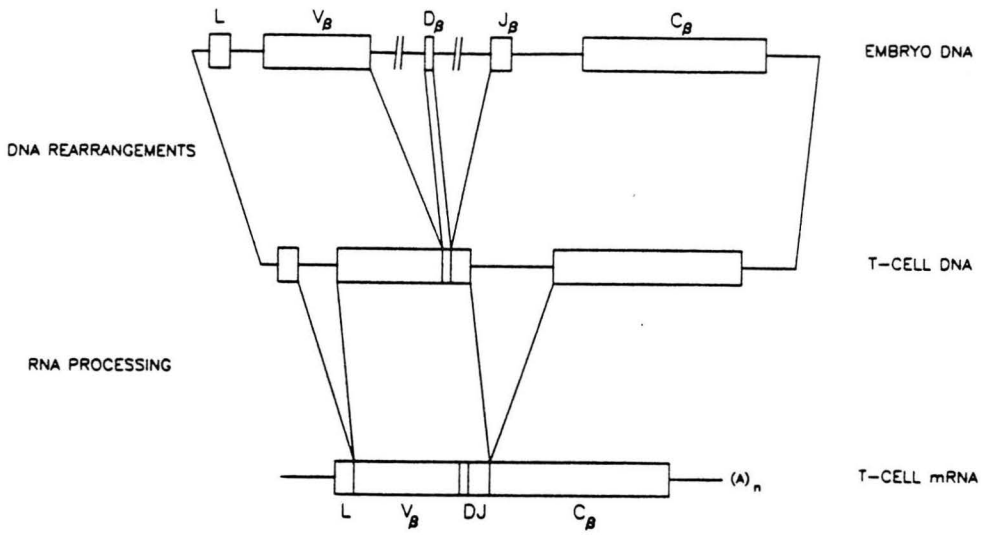


Figure 7. A Diagrammatic Representation of the Gene Segments and DNA Rearrangements and RNA Processing Events Necessary for the Expression of the β mRNA of the T Cell Antigen Receptor

The C_β gene has three introns which are not indicated (Malissen, unpublished observations).

to generate a contiguous V_β gene (Figure 7). In addition, the β polypeptide is an integral membrane protein like its immunoglobulin heavy chain counterpart and in contrast to the immunoglobulin λ and κ chains. The fact that the V_β gene has three genetic elements that presumably can rearrange in a combinatorial manner provides the V_β gene with combinatorial diversification potential equal to that of V_H genes and greater than that of V_L genes. In addition, V_β genes may have still greater diversificiation potential through D_β–D_β or V_β–J_β joining alternatives. In this regard, it will be interesting to determine whether the T cell receptor α genes also employ three rearranging elements.

Experimental Procedures

Isolation of Germ-Line Clones

Genomic libraries of human sperm DNA and of DNA from the human osteosarcoma U2 OS were constructed by A. Winoto and L. Fors as previously described (Steinmetz et al., 1982). These libraries were screened with the complete YT35 cDNA probe (Yanagi et al., 1984) using the procedure of Steinmetz et al. (1982).

Isolation of cDNA Clones

A cDNA library from mRNA isolated from the Jurkat tumor line was constructed and cloned into the Pst I site of pBR322 and screened with the YT35 cDNA probe according to the protocol of Yanagi et al. (1984).

Genomic Blot Hybridizations

Ten micrograms of DNA from human sperm and 10 μg from a B cell line from the clone FQ were digested with restriction enzymes and subjected to electrophoresis on horizontal agarose slab gels. After denaturation and neutralization, the DNA was transferred to nitrocellulose paper. Hybridizations were carried out at 68°C in a solution of 1 M NaCl, 0.05 M Tris-HCl (pH 7.5), 0.1% NaPPi, 0.1% SDS, 10× Denhardt's solution, 150 μg/ml salmon sperm DNA, and 0.5 μg nick-translated ^{32}P-labeled YT35 probe (1.5–2 × 10^8 dpm/μg). Prehybridization of the filter in hybridization buffer was carried out for 3 hr at 68°C. Hybridizations were carried out for 15 to 20 hr at 68°C. Following hybridization, the filters were washed three times in high-salt wash buffer (3× SSC, 0.1% SDS) and three times in low-salt wash buffer (1× SSC, 0.1% SDS). Each wash was carried out for 30 min at 68°C. Filters were exposed for 24 hr at −70°C with an intensifying screen.

DNA Subcloning and Restriction Mapping

Restriction fragments containing the V and J genes were subcloned into the plasmid pUC8 (Vieira and Messing, 1982) and further subcloned into phages M13mp8 or M13mp9 (Messing and Vieira, 1982). Restriction map analyses were carried out using single and double digests of infrequent-cutting enzymes on the subclones.

DNA Sequencing

The $J_{\beta 1}$ fragment was sequenced according to the method of Maxam and Gilbert (1980). The V region sequencing was carried out using the Sanger method (Sanger et al., 1977) as modified by E. Strauss (unpublished results). Dideoxy sequencing in the M13mp8 cloning vector was carried out using both the standard primer and specific oligonucleotide primers synthesized by Dr. S. Horvath (Horvath et al., submitted). All DNA sequences were sequenced on both strands.

Acknowledgments

The authors thank Drs. E. Reinherz, J. Kappler, and P. Marrack for making available experimental results prior to publication, Dr. Suzanna Horvath for the synthetic oligonucleotides, Lance Fors and Dr. Winoto for the cosmid libraries, and Tim Hunkapiller and Drs. Joan Kobori, Mitchell Kronenberg, and Roger Perlmutter for helpful advice and discussions and for critical reading of the manuscript. The authors would also like to thank Debbie Maloney for excellent technical assistance, the people at the Graphic Arts facility and the photo labs for excellent, efficient work, and the secretarial staff of the Division of Biology at California Institute of Technology for patient preparation of the manuscript.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 13, 1984

References

Acuto, O., Hussey, R. E., Fitzgerald, K. A., Protentis, J. P., Meuer, S. C., Schlossman, S. F., and Reinherz, E. L. (1983). The human T cell receptor: appearance in ontogeny and biochemical relationship of α and β subunits on IL-2 dependent clones and T cell tumors. Cell *34*, 717–726.

Acuto, O., Fabbi, M., Smart, J., Poole, C., Protentis, J., Royer, H., Schlossman, S., and Reinherz, E. (1984). Purification and N-terminal amino acid sequencing of the β subunit of a human T-cell antigen receptor. Proc. Nat. Acad. Sci. USA, in press.

Alt, F., and Baltimore, D. (1982). Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D–J_H fusions. Proc. Nat. Acad. Sci. USA *79*, 4118–4122.

Baltimore, D. (1974). Is terminal deoxynucleotidyl transferase a somatic mutagen in lymphocytes? Nature *248*, 409–411.

Bentley, D. L., and Rabbits, T. H. (1980). Human immunoglobulin variable region genes—DNA sequences of two V_κ genes and a pseudogene. Nature *288*, 730–733.

Bentley, D. L., and Rabbits, T. H. (1983). Evolution of immunoglobulin V genes: evidence indicating that recently duplicated human V_κ sequences have diverged by gene conversion. Cell *32*, 181–189.

Blomberg, B., and Tonegawa, S. (1982). DNA sequences of the joining regions of mouse λ chain immunoglobulin genes. Proc. Nat. Acad. Sci. USA *79*, 530–533.

Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, F., Rajewsky, K., and Baltimore, D. (1981). Heavy chain variable region contribution to the NP^b family of antibodies: somatic mutation evident in a γ2A variable region. Cell *24*, 625–637.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. *50*, 349–383.

Crews, S., Griffin, J., Huang, H., Calame, K., and Hood, L. (1981). A single V_H gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of the antibody. Cell *25*, 59–66.

Early, P., Huang, H., Davis, M., Calame, K., and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. Cell *19*, 981–992.

Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D., and Cohen, J. (1981). Diversity of germline immunoglobulin V_H genes. Nature *292*, 426–430.

Gough, N., and Bernard, O. (1981). Sequences of the joining region genes for immunoglobulin heavy chains and their role in generation of antibody diversity. Proc. Nat. Acad. Sci. USA *78*, 509–513.

Hedrick, S., Cohen, D., Nielsen, E., and Davis, M. (1984a). Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. Nature *308*, 149–153.

Hedrick, S., Nielsen, E., Kavaler, J., Cohen, D., and Davis, M. (1984b). Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. Nature *308*, 153–158.

Heiter, P., Maizel, J., and Leder, P. (1982). Evolution of human immunoglobulin κ J region genes. J. Biol. Chem. *257*, 1516–1522.

Honjo, T. (1983). Immunoglobulin genes. Ann. Rev. Immunol. *1*, 499–528.

Jerne, N. K. (1974). Towards a network theory of the immune system. Ann. Immunol. (Inst. Pasteur) *125C*, 373–389.

Kappler, J., Kubo, R., Haskins, K., White, J., and Marrack, P. (1983). The mouse T cell receptor: comparison of MHC-restricted receptors on two T cell hybridomas. Cell *34*, 727–737.

Kim, S., Davis, M., Sinn, E., Patten, P., and Hood, L. (1981). Antibody diversity: somatic hypermutation of rearranged V_H genes. Cell *27*, 573–581.

Kraig, E., Kronenberg, M., Kapp, J., Pierce, C., Abruzzini, A., Sorensen, C., Samelson, L., Schwartz, R., and Hood, L. (1983). T and B cells that recognize the same antigen do not transcribe similar heavy chain variable region gene segments. J. Exp. Med. *158*, 192–209.

Kronenberg, M., Kraig, E., and Hood, L. (1983a). Finding the T-cell antigen receptor: past attempts and future promise. Cell *34*, 327–329.

Kronenberg, M., Kraig, E., Siu, G., Kapp, J., Kappler, J., Marrack, P., Pierce, C., and Hood, L. (1983b). Three T-cell hybridomas do not contain detectable heavy chain variable gene transcripts. J. Exp. Med. *158*, 210–227.

Kurosawa, Y., and Tonegawa, S. (1982). Organization, structure, and assembly of immunoglobulin heavy chain diversity DNA sequences. J. Exp. Med. *155*, 201–218.

Kurosawa, Y., von Boehmer, H., Haas, W., Sakano, H., Trauneker, A., and Tonegawa, S. (1981). Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. Nature *290*, 565–570.

Loh, D. Y., Bothwell, A. L. M., White-Scharf, M. E., Imanishi-Kari, T., and Baltimore, D. (1983). Molecular basis of a mouse strain-specific anti-hapten response. Cell *33*, 85–93.

Litman, G., Berger, L., Murphy, K., Litman, R., Hinds, K., Jahn, C., and Erickson, B. (1983). Complete nucleotide sequence of an immunoglobulin V_H gene homologue from Caiman, a phylogenetically ancient reptile. Nature *303*, 349–352.

Max, E., Maizel, J., and Leder, P. (1981). The nucleotide sequence of a 5.5-kilobase DNA segment containing the mouse κ immunoglobulin J and C region genes. J. Biol. Chem. *256*, 5116–5120.

Maxam, A., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. Meth. Enzymol. *65*, 499–560.

McIntyre, B. W., and Allison, J. P. (1983). The mouse T cell receptor: structural heterogeneity of molecules of normal T cells defined by xenoantiserum. Cell *34*, 739–746.

Messing, J., and Vieira, J. (1982). A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. Gene *19*, 269–276.

Otto, R., and Rougeon, F. (1983). Gene conversion and polymorphism: generation of mouse immunoglobulin γ2a chain alleles by differential gene conversion by γ2b chain gene. Cell *32*, 515–523.

Perlmutter, R., Klotz, J., Bond, M., Nahm, M., Davie, J., and Hood, L. (1984). Multiple V_H gene segments encode murine antistreptococcal antibodies. J. Exp. Med. *159*, 179–192.

Rechavi, G., Bienz, B., Ram, D., Ben-Neriah, Y., Cohen, J., Zakut, R., and Givol, D. (1982). The organization and evolution of immunoglobulin V_H gene subgroups. Proc. Nat. Acad. Sci. USA *79*, 4405–4409.

Samuelson, L., Germain, R., and Schwartz, R. (1983). Monoclonal antibodies against the antigen receptor on a cloned T-cell hybrid. Proc. Nat. Acad. Sci. USA *80*, 6972–6976.

Sanger, F., Nicklen, S., and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Nat. Acad. Sci. USA *74*, 5463–5467.

Seidman, J., Leder, A., Nau, M., Norman, B., and Leder, P. (1978). Antibody diversity. Science *202*, 11–17.

Steinmetz, M., Winoto, A., Minard, K., and Hood, L. (1982). Clusters of genes encoding mouse transplantation antigens. Cell *28*, 489–498.

Tonegawa, S. (1983). Somatic generation of antibody diversity. Nature *302*, 575–581.

Vieira, J., and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene *19*, 259–268.

Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S., Aleksander, I., and Mak, T. (1984). A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. Nature *308*, 145–149.

CHAPTER FOUR

ANALYSIS OF A HUMAN V_β GENE SUBFAMILY

Submitted for publication to the *Journal of Experimental Medicine*

ANALYSIS OF A HUMAN V_β GENE SUBFAMILY

By GERALD SIU, ERICH C. STRAUSS, and LEROY E. HOOD

From the Division of Biology, California Institute of Technology

Pasadena, California 91125

Running Title: HUMAN V_β SUBFAMILY

The T-cell antigen receptor consists of two chains, denoted α and β, each of which is composed of a variable (V) region and a constant (C) region. The genes that encode both the α and β chains show extensive homology to the genes that encode the B-cell receptor, the immunoglobulin (1, 2). Each is encoded by two genes: a V gene of approximately 360 bp and a C gene of approximately 500 bp (3, 4). The gene that encodes the V region of the β chain consists of three gene segments, V_β, D_β, and J_β, that are joined together during T-cell development to generate the complete V_β gene (5, 6). The V_β gene segment encodes the first 280-300 bp, the D_β gene segment the next 5-10 bp, and the J_β gene segment the final 45-51 bp of the V_β gene (5-11). Recognition signals that mediate immunoglobulin V gene rearrangement appear to mediate V_β gene rearrangement as well. These signals consist of conserved heptamer and nonamer sequences that are separated by a variable spacer sequence of either 12 bp or 23 bp (5, 6).

Immunoglobulin V gene segments are grouped into subfamilies of closely related gene segments. Each subfamily has from four to over 50 members, each member 75% or more similar at the DNA level to the other members (12-14). In contrast, the V_β gene segments in mouse appear to consist of many single or low copy subfamilies. Of the V_β gene segment subfamilies analyzed in mouse to date, twelve contain only one member, and two contain three members (15-17).

Analyses of the V_β coding regions have revealed that the V_β genes appear to be very similar to immunoglobulin V genes. Hypervariability plots of the V_β genes reveal patterns not unlike those of the V_H and V_L genes; analyses of the beta-strand-forming potential and the relative hydrophobicity of the V_β regions have also shown them to be very similar to immunoglobulins (15-17). We were interested in analyzing the coding and flanking regions of the germline V_β gene segments in order to see if these gene segments resemble immunoglobulin V genes in their germline structure. In addition, we were interested in studying the

germline organization of the V_β gene family. We have used as a model system the $V_{\beta M3}$ subfamily, a human V_β gene subfamily of five members that represents the largest V_β subfamily studied to date. The members of this subfamily were identified by their similarity to the V_β region of the YT35 cDNA (3). We have isolated cosmids and lambda clones containing the five members of this family and each was further characterized by subcloning and sequence analysis.

Materials and Methods

Genomic Blots. Genomic blots were carried out as previously described (18, 19). DNA probes for the hybridizations were labelled with ^{32}P utilizing the protocol of Rigby et al. (19). Hybridizations were carried out at 68°C for 24 h. After hybridization, the filters were washed three times with 3X SSC for 20 min each followed by three washes with 1X SSC for 20 min each, at 68°C. The filters were then exposed to Kodak XAR-5 film overnight at -70°C with an intensifying screen.

Construction and Screening of Genomic Libraries. A human sperm cosmid library was constructed by Lance Fors (unpublished), and screened as previously described (19). Partial lambda libraries were constructed into the vector λgt7lac5 and screened as described (21).

Restriction-endonuclease Mapping. Mapping of different restriction endonuclease sites was accomplished using single- and double-digests of different enzymes and comparing them to each other. In addition, a novel restriction-endonuclease-mapping technique (Y. H. Sun and L. E. Hood, in preparation) based on the technique of Smith and Birnsteil (22) was utilized. This technique depends on the presence of a Sal I, Cla I or Nru I site present in the cloning vector and not in the insert DNA. In our case, the cosmid was digested with Sal I to completion, phenol-extracted, ethanol precipitated, and resuspended to a final concentration

of 100 ng/μl. 200 ng aliquots of the cosmid were then digested with different concentrations of a certain enzyme, that would permit partial, but not complete digestion. A range of the partial-digestion reactions was taken and pooled, and 80 ng of the sample was then subjected to electrophoresis on two sets of 0.6% agarose horizontal slab gels, the first for 300-400 volt-hours, the second for 1.8-2X as long. The gels were run in an electric field of 1.7 volts/cm to permit good resolution. After electrophoresis, the gels were denatured, neutralized, blotted with nitrocellulose and the nitrocellulose baked as described above. Two synthetic oligonucleotide hybridization probes corresponding to the regions of DNA on either side of the Sal I site in the cosmid vector were used for screening. The oligonucleotides were labelled with ^{32}P using polynucleotide kinase according to the protocol of Maxam and Gilbert (23), and hybridized separately to each set of filters and the filters washed three times in 3X SSC at 68°C for 15 min each. The filters were then exposed to x-ray film overnight. The probes will hybridize to each partial-digestion product that contains the vector sequence adjacent to the Sal I site on one end and a recognition site for the restriction enzyme tested on the other end. The hybridization will result in a ladder of bands, each corresponding to a different restriction enzyme site in the clone. In this manner, the restriction sites for each restriction enzyme were mapped with respect to the Sal I site of the vector. Both oligonucleotides were necessary in order to accurately map all of the restriction sites in the clones.

Subcloning and DNA Sequencing. Subcloning was carried out using the procedures described previously (19). Sequencing was carried out using the dideoxynucleotide sequencing technique as described by Strauss et al (24).

Results

Isolation of genomic V_β Clones. A V region-specific probe was isolated

from the YT35 cDNA clone. With this probe, five hybridizing bands were detected on a genomic blot of human DNA at high wash stringencies (Fig. 1). These five bands corresponded to five different V_β gene segments, which were called $V_{\beta M3-0}$, $V_{\beta M3-1}$, $V_{\beta M3-2}$, $V_{\beta M3-3}$, and $V_{\beta M3-4}$. A cosmid containing the $V_{\beta M3-1}$ and $V_{\beta M3-2}$ genes was isolated; the $V_{\beta M3-2}$ gene segment was subcloned and characterized previously (6). From the same human cosmid library, two additional overlapping cosmids were isolated that contained the $V_{\beta M3-4}$ gene segment. In order to isolate the remaining members, partial libraries were constructed using the λgt7lac5 vector from DNA from the same individual and screened using the V-region-specific probe. Figure 2 shows the restriction maps of the three cosmids and the two lambda clones that contain all five members of the $V_{\beta M3}$ family. The cosmid H7.1 contains the $V_{\beta M3-1}$ and $V_{\beta M3-2}$ gene segments, the cosmids H9.1 and H18.1 contain the $V_{\beta M3-4}$ gene segment, the lambda clone λBM3-0 contains the $V_{\beta M3-0}$ gene segment, and the lambda clone λBM3-3 contains the $V_{\beta M3-3}$ gene segment. A total of 38 kb around the $V_{\beta M3-1}$-$V_{\beta M3-2}$ cluster, 40 kb around $V_{\beta M3-4}$, 9 kb around $V_{\beta M3-0}$, and 4.3 kb around $V_{\beta M3-3}$ have been cloned. In order to determine if additional V_β gene segments homologous to the YT35 V-specific probe are present on these clones, the cosmid and lambda clones were digested with restriction enzymes and blotted to nitrocellulose filters filters were then hybridized to the V region-specific probe and washed under low stringency conditions. No hybridization of the V region-specific probe was observed in the flanking regions (data not shown). The five members of the $V_{\beta M3}$ subfamily were identified by restriction mapping, subcloned, and the coding and flanking regions sequenced and analyzed (Fig. 3).

The $V_{\beta M3-0}$ gene segment is 77.8% homologous to the V region-specific probe at the DNA level. The coding region contains an in-frame stop codon at amino acid position 51 (Fig. 4); it is probable, therefore, that $V_{\beta M3-0}$ is a

pseudogene. The $V_{\beta M3-1}$ gene segment is identical to the hybridization probe. The $V_{\beta M3-2}$ gene segment was isolated and characterized earlier and is located 3 kb 3' to the $V_{\beta M3-1}$ gene segment. The $V_{\beta M3-3}$ gene segment is 82.1% and the $V_{\beta M3-4}$ gene segment is 79.2% similar to the hybridization probe. The $V_{\beta M3-4}$ coding region contains an in-frame stop codon at amino acid position 64 (Fig. 4). Like $V_{\beta M3-0}$, this gene segment is probably a pseudogene.

Each of the members of the $V_{\beta M3}$ family has all of the characteristics of a germline V_β gene segment. The $V_{\beta M3}$ gene segments consist of two exons, one of 46 bp that encodes most of the leader sequence, and the other of approximately 297 bp that encodes the remainder of the leader as well as the variable region. These exons are split by an intervening sequence of approximately 100 bp; this is similar in size to immunoglobulin V gene introns, although the identification of larger introns in other V_β gene segments indicates that the size of the intron is variable. Comparison of the translated germline sequences with the N-terminal protein sequence of a β-chain utilizing $V_{\beta M3-1}$ from the REX tumor indicates that the N-terminal-most residue is the glycine located 20 amino acids from the conserved cysteine (Fig. 4) (25). The intron, then, divides the leader at codon position -5 (Fig. 3). This differs from immunoglobulins, where the intron splits the leader at the codon position -4 for both light and heavy chains.

Sequence Comparisons. The sequences of the five gene segments and flanking regions were subdivided into five functionally distinct regions and compared with one another (Table I). The $V_{\beta M3-1}$ and $V_{\beta M3-2}$ gene segments are the most similar in both coding and flanking regions. In all, these two gene segments exhibit 93.9% and 97.7% sequence similarity in the two coding exons, and 94.3%, 95.2%, and 97% in the 5' and 3' flanking regions and the intron, respectively (Table I). Additional sequence analysis has revealed that this similarity extends for a total of at least 1.5 kb in the 5' and 3' flanking regions

(data not shown). Comparisons of the other $V_{\beta M3}$ gene segments to $V_{\beta M3-1}$ and $V_{\beta M3-2}$ indicate that they are significantly less similar. In general, the coding region appears to be the most conserved and the flanking and intron regions less so. $V_{\beta M3-3}$ is similar to the others only in the coding regions and in the immediate 3' flanking region. With the exception of a conserved 16 bp sequence, there appears to be little or no sequence similarity in the 5' flanking region. $V_{\beta M3-0}$ is fairly similar in both the coding and the flanking regions to $V_{\beta M3-1}$, $V_{\beta M3-2}$ and $V_{\beta M3-4}$. The 5' flanking region of $V_{\beta M3-0}$ contains a 27 bp direct repeat of a sequence that is found in one copy in the flanking regions of $V_{\beta M3-1}$, $V_{\beta M3-2}$ and $V_{\beta M3-4}$. The repeats are located adjacent to each other, separated by one base pair (Fig. 3).

Mutation frequencies. As mentioned above, the coding regions appear to be more conserved than the flanking and intron sequences. This would indicate that selection pressure that acts to maintain the coding region sequence may exist. In order to study this possibility, we have analyzed the mutation frequencies and rates for the coding regions of the members of the $V_{\beta M3}$ family using the method of Kimura (26) (Table II). In this method, the number of mutation events that occurred between two similar sequences are counted and separated into different groups. In one case, the mutations that resulted in replacements in the amino acid sequence (replacement-site mutations) and those that did not result in replacements (silent-site mutations) were separated. In the other case, the mutations in the different parts of each codon were separated; thus, all the mutations that occurred in the first, second and third position of each codon were tabulated separately. The mutation frequencies were then calculated using the Kimura equation, and the mutation rates were calculated using the equation

$$K = 2kT$$

where K is the mutation frequency, k is the mutation rate, and T is the divergence time (27). In each comparison, the mutation frequency in the replacement-site is much lower than the frequency in the silent site. The ratio of replacement-site mutation frequency over the silent site mutation frequency, K_A/K_S, ranges from 0.19 to 0.29 for the functional genes, which is comparable to that of other eukaryotic genes. The K_A/K_S ratio for the comparisons with the pseudogenes, however, ranges from 0.33 to 0.41.

The Human V_β Gene Family. Comparison of the $V_{\beta M3}$ subfamily with other human V_β genes indicates that, as in the murine system, the V_β gene family is quite diverse (Fig. 4). The most divergent human V_β genes, MOLT4 (28) and HPB-ALL (29), differ from each other by 67% on the DNA level and 74% on the protein level (Table III). Interestingly, the V_β gene segment utilized by the HPB-MLT tumor (30) exhibits sequence similarity to the $V_{\beta M3}$ family. This V_β gene differs by only 29% and 39% on the DNA and the protein levels, respectively, to $V_{\beta M3\text{-}1}$. Because of the lesser similarity, this V_β gene segment is not detectable on a genomic blot under high wash stringencies, and therefore was not detected earlier.

Discussion

The $V_{\beta M3}$ Subfamily. The human $V_{\beta M3}$ subfamily consists of five members that were identified by hybridization of a V region-specific probe from the YT35 cDNA to a genomic blot of human DNA. The $V_{\beta M3}$ subfamily is larger than any of the murine V_β gene subfamilies characterized to date and is the size of some immunoglobulin V gene families. A monoclonal antibody that binds to V_β regions encoded by the gene segments of this subfamily was shown to bind to 2% of human peripheral blood T lymphocytes, and was found to bind to T cells of both the $T4^+$ and $T8^+$ phenotypes as well as separate T_H, T_K, and T_S populations, indicating

that some of the gene segments in this subfamily can be used in all classes of T lymphocytes (31). Interestingly, a T_H clone identified by this antibody as using a $V_{\beta M3}$ family member (in the nomenclature of Acuto *et al.* (31) V_βREX) has subsequently been found to rearrange the 2.0 kb Eco RI $V_{\beta M3}$-containing restriction fragment (Fig. 1). Thus, this T_H clone probably utilizes $V_{\beta M3\text{-}2}$ in the β chain of its T-cell receptor. In addition, the $V_{\beta M3\text{-}1}$ gene segment is identical to the V_β regions of the YT35 and JM cDNAs, indicating that $V_{\beta M3\text{-}1}$ is the germline gene segment used in the β-chain of the T-cell receptor of the MOLT-3 (3) and JM (32) tumors. This also indicates that somatic hypermutation has not occurred in the beta-chain gene in these tumors. This is consistent with other studies, in which comparisons between the germline gene segments and the rearranged V_β genes have shown no evidence for somatic hypermutation and implies that somatic hypermutation does not occur in the β chain (5, 33).

Structure and Diversity of Human V_β Gene Segments. Analyses of the $V_{\beta M3}$ subfamily members and comparisons with other human Vβ gene segments as well as all the other V gene segment family members have identified many amino acids that are conserved both sequence and position in all V gene segment coding regions. Previous structural analyses of immunoglobulins have identified conserved interactions that are believed to be important in the structure of the variable region; in most cases these interactions appear to involve these conserved amino acids (34-36). Of the five conserved amino acids that are believed to be important for the interaction of the V_H and V_L regions, four are conserved in the V_β region. These include the tyr 34, the gln 36, the leu 45 (Fig. 4) and a phenylalanine in the J_β region (not shown). The tyr 34 and the phe in the J gene segment are conserved in light chains, the leu 45 is conserved in heavy chains, and the gln 36 is conserved in both. Other conserved amino acids that are believed to be important for V_H or V_L the stabilization of the domains are conserved in the V_β region as well. These

amino acids include glu 4, pro 6, gly 16, cys 21, trp 33, leu 77, tyr 90, cys 92, and ala 93. Analyses of the V_α regions have shown that most of these conserved amino acids are also present in V_α regions. Taken together, these data imply that a V_α-V_β heterodimer will be similar in tertiary structure to the immunoglobulin V_H-V_L binding site.

Despite these conserved amino acids, the human Vβ gene segments can differ from each other considerably. Comparison of the $V_{\beta M3}$ subfamily members with other human V_β gene sequences show that the human V_β gene segments can differ by as much as 67% and 74% on the DNA and protein levels, respectively. This is comparable to the diversity observed in the murine V_β gene family, where gene segments can differ by 62% and 72% on the DNA and the protein levels, respectively (15-17). This indicates that the range of diversity in germline human V_β gene segments is equivalent to that of their murine counterparts. Interestingly, the V_β gene segment utilized by the HPB-MLT tumor (30) shares significant homology with the members of the $V_{\beta M3}$ family. This V_β gene segment shares the most homology with the $V_{\beta M3-1}$ gene segment to which it is 71% and 61% homologous on the DNA and protein levels, respectively. These data indicate that, like the immunoglobulin V gene families and the murine V_α gene family, the human V_β gene family may consist of multimember gene subfamilies that are somewhat homologous to each other. This is supportive of the theory that the V gene families were created by successive duplication, generated by mechanisms such as unequal crossing-over events (37). This theory predicts that the V genes will exhibit a continuum of homology, ranging from recently-duplicated and therefore closely-related genes, to genes that diverged so long ago that they retain only the homology imposed by V region structure. This is clearly seen in the human V_β gene family. In contrast, the murine V_β gene family consists of mostly single-member subfamilies that share little homology to each other (15-

17). The significance of this difference, if any, is unknown.

Analyses of the DNA sequences of the flanking regions of the members of the $V_{\beta M3}$ subfamily have identified several sequences that resemble eukaryotic sequences. Eukaryotic promoter sequences generally include an A/T rich region with a canonical ATA sequence called the TATA sequence that is usually found 21-23 bp 5' to the initiation point of transcription, and the sequence CCAAT that is found 5' to the TATA sequence (38). For the $V_{\beta M3-1}$ and $V_{\beta M3-2}$ gene segments, TATA-like sequences are located 25 bp and 89 bp 5' to the methionine codon of the leader. This may indicate that there are multiple initiation points for transcription for these gene segments. The $V_{\beta M3-4}$ gene segment has only the position 89 TATA-like sequence, whereas the $V_{\beta M3-0}$ gene segment does not appear to have the canonical ATA sequence in its 5' flanking region although an A/T-rich region is present in a similar position as the TATA sequences of the $V_{\beta M3-1}$, $V_{\beta M3-2}$ and $V_{\beta M3-4}$ gene segments. In addition, a sequence that somewhat resembles the canonical CCAAT sequence is located 34 bp 5' to the TATA sequence (Fig. 3).

Analyses of the 5' flanking regions of other germline T-cell receptor V gene segments show that many appear to have similar promoter-like sequences. The murine $V_{\beta 1}$ (39) gene segment and the V_{γ} (40) gene segments have the consensus TATA sequences. Interestingly, these gene segments also appear to have the same CCAAT-like sequence as the $V_{\beta M3-1}$, $V_{\beta M3-2}$, $V_{\beta M3-4}$ and $V_{\beta M3-0}$ gene segments. This sequence, TGGCCCATTC, includes additional sequences that are not part of the classically-defined CCAAT sequence. Other murine V_{β} gene segments, however, do not appear to have promoter sequences in the immediate 5' flanking regions. Both the murine $V_{\beta 2}$ and $V_{\beta 14}$ gene segments do not appear to have either TATA sequences or this additional sequence, and yet are expressed in functional T-cell lines (M. Malissen, unpublished data). It is possible that the 5'

untranslated regions of these gene segments are longer, or an intron exists within this region. As with the $V_{\beta 2}$ and $V_{\beta 14}$ gene segments, the $V_{\beta M3-3}$ gene segment does not have promoter-like sequences; the only conserved portion between $V_{\beta M3-3}$ and the other members of the $V_{\beta M3}$ subfamily is a 16 bp sequence located 90 bp 5' to the initiation codon, and 15 bp 5' to the TATA-like sequence. Its significance, if any, is unknown. Immunoglobulin V gene segments have an additional region of transcription control located just 5' to the TATA sequence, known as the octamer sequence; none of the V_{β} gene segment flanking regions, however, has these sequences. Additional data are necessary to conclusively show if T-cell receptor-specific promoter sequences exist.

Analysis of the $V_{\beta M3-3}$ gene segment indicates that the V coding region and the immediate 3' flanking region appears to be far more similar to the homologous regions in the other $V_{\beta M3}$ subfamily members than the intron and 5' flanking region (Table I). Although this comparison includes the V coding region, which is probably under selection pressure (see below), comparable numbers are observed if only noncoding sequences are considered. The reasons for this asymmetry in similarity are not known, as both the 5' and 3' flanking regions do not represent coding regions and therefore should not be subjected to selection pressure at that level. The similarity appears to begin at the middle of the intron (position 410 in Fig. 3) and end 110 bp 3' to the gene segment. The similarity within this region between $V_{\beta M3}$ subfamily ranges from 71.2 to 77%, while the similarity in the other regions ranges from 46.3 to 50.9% (Table I and data not shown). Although the heptames and nonamer sequences of the rearrangement recognition signals are present in the immediate 3' flanking region, these sequences are too small to significantly alter these calculations. It is tempting to speculate that this asymmetry is the result of a gene conversion event that corrected the immediate 3' part of the intron, the V coding region, and the 3'

flanking region of the $V_{\beta M3-3}$ gene segment. Similar gene conversion events have been proposed in the immunoglobulin V gene families (44).

The 3' flanking region contains the recognition signals for DNA rearrangement necessary for the variable region gene formation. As with all V_β gene segments the $V_{\beta M3}$ subfamily members have recognition signals with spacer lengths of 23 base pairs (Fig. 3).

V_β Gene Evolution. Analysis of the members of the $V_{\beta M3}$ family indicates that the coding region is the most highly conserved. The $V_{\beta M3-1}$ and $V_{\beta M3-2}$ gene segments are the most similar, indicating that one arose from the other during a recent gene duplication event. The others are approximately equidistant from each other in sequence similarity. Analysis of the silent-site mutation frequency indicates that the members of the $V_{\beta M3}$ family, with the exception of $V_{\beta M3-1}$ and $V_{\beta M3-2}$, diverged from each other greater than 100 million years ago, prior to the mammalian divergence. This calculation assumes that the silent-site mutation rate is 5.1×10^{-9}/site•yr, which is believed to be approximately the same for all genes (27). Analysis of the mutation frequencies in the replacement sites show that the replacement-site mutation frequency is much lower than the silent-site frequency. The ratio of replacement-site mutation frequency to silent-site mutation frequency provides a method in which to quantitate the rate at which two homologous genes are diverging from each other (27). This ratio ranges from 0.19 to 0.41, but from only 0.19 to 0.29 if comparisons of the functional gene segments are considered. This is much lower than that found in immunoglobulin V_H gene segments (G.S., S. Crews, E. Springer, H. Huang, and L.E.H., in preparation), and approximately equivalent to that of most other eukaryotic genes (27). In comparing the mutation frequencies in the three positions of the V_β codons, it is apparent that the lowest mutation frequency is in the second position, and the highest mutation frequency is in the third position. This is similar to

other genes, and is due to the fact that the second position of the codon has no silent sites, and the third position of the codon has the most silent sites (45). In addition, mutations in the second position are more likely to result in nonconservative replacements; that is, amino acid replacement changes in the second position result in greater changes in the physical properties of the protein (45). These data also contrast with the V_H gene segments; there, the lowest frequency is in the first position, and the second position is consistently higher than the third position (G.S., S. Crews, E. Springer, H. Huang, and L.E.H., in preparation).

A murine V_β gene segment homologous to the $V_{\beta M3}$ subfamily was recently identified in a β-chain cDNA isolated from a mouse spleen cDNA library (17). Unlike the human subfamily, the murine $V_{\beta M3}$ subfamily is a single-gene subfamily (17). This gene segment, denoted $V_{\beta 11}$ (2), is most closely related to $V_{\beta M3\text{-}2}$ (Table II). The mutation frequencies in the amino-acid-replacement- and silent-sites and the three codon positions were calculated between these two gene segments (Table I). Like the other comparisons between the functional members of the $V_{\beta M3}$ family, the K_A/K_S ratio and the relative values of K_1, K_2 and K_3 are similar to other eukaryotic genes, indicating that selection pressure is acting to maintain the coding region sequence. Using the K_A and assuming that humans and mice diverged at mammalian radiation (85 million years ago), the mutation rate in the amino-acid replacement site is 1×10^{-9}/site•yr, which is comparable to the mutation rates in β-globin genes, 1.13×10^{-9}/site•yr (46), and less than that of IFN-α, 2.1×10^{-9}/site•yr. Analyses of a large number of different eukaryotic genes have indicated that most have amino-acid replacement-site mutation rate in the range of $0.2\text{-}2.0 \times 10^{-9}$/site•yrs (48). Previously, studies involving murine V_β-specific hybridization probes showed that homologous V_β gene segments could not be detected in species that were closely related to the mouse, even though it was

possible to do so with murine immunoglobulin V probes. It was therefore proposed that V_β gene segments were mutating more quickly than immunoglobulin V gene segments. Our data, however, are inconsistent with this theory, and indicate that the V_β gene segments are diverging at the rate of most eukaryotic genes.

V_β Pseudogenes. Two of the five members of the $V_{\beta M3}$ family, $V_{\beta M3-0}$ and $V_{\beta M3-4}$, are pseudogenes in that they share significant homology with a functional gene, but have mutations that would prevent their expression. Both have in-frame stop codons that would prevent translation of a variable region gene that utilizes one of these gene segments. Interestingly, analyses of the mutation rates in the two pseudogenes show that although they are mutating faster than the functional members of this family, the difference is not statistically significant (Table I). This implies that both $V_{\beta M3-0}$ and $V_{\beta M3-4}$ have become pseudogenes fairly recently, and therefore have not yet accumulated many mutations in the replacement sites. Consistent with this is the observation that the in-frame stop codons are the only crippling mutations in these gene segments; pseudogenes usually acquire many crippling mutations over time since there is no selection pressure to inhibit this process. The somewhat slow mutation rate in the replacement site of pseudogenes is similar to what is observed in the immunoglobulin T15 V_H subfamily. In this case, a recently-created pseudogene does not appear to be mutating faster than the functional gene segments in this subfamily, even though the pseudogene has several crippling mutations (G.S., S. Crews, E. Springer, H. Huang, and L.E.H., in preparation.

The presence of two pseudogenes in a subfamily of five members indicates that the proportion of pseudogenes in the V_β family may be very high, perhaps as high as that of the immunoglobulin V gene families, in which perhaps 25-30% of the members are nonfunctional. In addition, V_β pseudogenes have been identified in β chain cDNAs isolated from murine and human thymus cDNA libraries (17, P.

Concannon, L. Pickering and L.E.H., in preparation). The high proportion of pseudogenes in the V gene families indicates that the generation of germline diversity results in a large number of nonfunctional gene segments.

Chromosomal Organization of V_β Gene Segments. Characterization of the genomic clones containing the five members of the $V_{\beta M3}$ family indicates that two of the five members, $V_{\beta M3-1}$ and $V_{\beta M3-2}$, are only 3 kb apart. As mentioned above, these two gene segments are most homologous to each other, and are probably the result of a recent gene duplication event. The other members of the family, however, do not appear to be as closely linked. $V_{\beta M3-0}$ was isolated on a 9 kb Eco RI fragment that does not overlap with any of the other clones. $V_{\beta M3-3}$ was isolated on a 4.3 kb Eco RI fragment; it, too, does not overlap with any of the other clones. $V_{\beta M3-4}$ was isolated on two overlapping cosmids; there is a total of 35 kb of 5' flanking DNA, and 6 kb of 3' flanking DNA. The $V_{\beta M3-1}$-$V_{\beta M3-2}$ cluster was isolated on one cosmid along with 29 kb of 5' flanking DNA and 5 kb of 3' flanking DNA. Using the V-region probe, we were unable to detect additional, less-homologous V_β gene segments on these clones. Variable spacing distances are also consistent with what is observed for immunoglobulin V gene segments. An analysis of the T15 subfamily of V_H gene segments reveals that the average spacing distance is approximately 23 kb, and ranges from 16 kb to greater than 60 kb (S. Crews, E. Springer, and G.S., unpublished data). The significance of the larger spacer distances is unclear. This organization is different from that seen in the rRNA families, which are characterized by short spacers of similar length (49). It is believed that the high homology in the rRNA families is maintained by frequent gene conversion events that are mediated by the already-existing extensive homology. V gene segments are characterized by diversity, and the occurrence of gene correction events could be detrimental towards the maintenance of a diverse repertoire. It is possible that the large distance between

homologous V_β gene segments may tend to reduce the frequency of gene correction events in order to maintain V_β diversity. Alternatively, the large spacer distances may be the results of genetic events that either created the gene segments from a common ancestor or occurred subsequently.

Our characterization of the $V_{\beta M3}$ subfamily reveals that the structure and organization of the human V_β locus is extremely similar to that of the immunoglobulin V gene loci. These similarities underscore the close structural and evolutionary relationships between immunoglobulins and the T-cell antigen receptor.

Summary

We have isolated five germline V_β gene segments that are homologous to the V region of the YT35 cDNA encoding the β-chain of the T-cell antigen receptor of the tumor MOLT-3. One of these gene segments is identical to the YT35 V region and therefore is the germline V_β gene segment that in part encodes the variable region of the YT35 cDNA. The other four members range from 77-98% similar to the YT35 V region. Analyses of the coding region sequences reveal that although the V_β gene segments are very diverse, they are mutating at a rate comparable to that observed in most eukaryotic genes. Analyses of the genomic clones show that the spacing distance between germline V_β gene segments ranges from 3 kb to greater than 30 kb.

Acknowledgments

We would like to thank Dr. Suzanna Horvath and Marilyn Tomich for providing synthetic oligonucleotide primers. We would also like to thank Drs. Richard Barth, Patrick Concannon, Joan Kobori, and Mitchell Kronenberg for helpful discussions and critique of the manuscript. Finally we would like to thank

Connie Katz and the other members of the secretarial staff at the Division of Biology at Caltech for preparation of the manuscript and the arrow-through-the-head.

References

1. Davis, M. M., Y.-H. Chien, N.R.J. Gascoigne, and S. M. Hedrick. 1984. A murine T-cell receptor gene complex: Isolation, structure and rearrangement. *Immunol. Rev.* **81**:235-258.

2. Kronenberg, M., G. Siu, L. Hood, and N. Shastri. 1986. The molecular genetics of the T-cell antigen receptor and T-cell antigen recognition. *Ann. Rev. Immunol.*, in press.

3. Yanagi, Y., Y. Yoshikai, K. Leggett, S. P. Clark, I. Aleksander, and T. W. Mak. 1984. A human T-cell specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* **308**:145-149.

4. Hedrick, S. M., E. A. Nielsen, J. Kavaler, D. I. Cohen, and M. M. Davis. (1984). Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. *Nature* **308**:153-158.

5. Chien, Y.-H., N.R.J. Gascoigne, J. Kavaler, N. E. Lee, and M. M. Davis. 1984. Somatic recombination in a murine T-cell receptor gene. *Nature* **309**: 322-326.

6. Siu, G., S. Clark, Y. Yoshikai, M. Malissen, Y. Yanagi, E. Strauss, T. Mak, and L. Hood. 1984. The human T-cell antigen receptor is encoded by variable, diversity and joining gene segments that rearrange to generate a complete V gene. *Cell* **37**:393-401.

7. Clark, S. P., Y. Yoshikai, S. Taylor, G. Siu, L. Hood, and T. W. Mak. 1984. Identification of a diversity segment of the human T-cell receptor beta chain, and comparison to the analogous murine element. *Nature* **311**:387-389.

8. Kavaler, J., M. M. Davis, and Y.-H. Chien. 1984. Localization of a T-cell receptor diversity-region element. *Nature* **310**:421-423.

9. Siu, G., M. Kronenberg, E. Strauss, R. Haars, T. Mak, and L. Hood. 1984. The structure, rearrangement and expression of D_β gene segments of the murine T-cell antigen receptor. *Nature* **311**:344-350.

10. Gascoigne, N.R.J., Y.-H. Chien, D. M. Becker, J. Kavaler, and M. M. Davis. (1984). Genomic organization and sequence of T-cell receptor β-chain constant- and joining-region genes. *Nature* **310**, 387-391.

11. Malissen, M., K. Minard, S. Mjolsness, M. Kronenberg, J. Goverman, T. Hunkapiller, M. Prystowsky, Y. Yoshikai, F. Fitch, T. Mak, and L. Hood. 1984. Mouse T-cell antigen receptor structure and organization of constant and joining gene segments encoding the β polypeptide. *Cell* **37**:1101-1110.

12. Cory, S., B. Tyler, and J. Adams. 1981. Sets of immunoglobulin V_κ genes homologous to ten cloned V_κ sequences: Implications for the number of germline V_κ genes. *J. Mol. Applied Genet* **1**:103-116.

13. Brodeur, P. and R. Riblet. 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur. J. Immunol.* **14**:922-930.

14. Dildrop, R. 1984. A new classification of mouse V_H sequences. *Immunol. Today* **5**:85-86.

15. Patten, P., T. Yokota, J. Rothbard, Y.-H. Chien, K.-I. Arai, and M. Davis. 1984. Structure, expression and diversity of T-cell receptor β-chain variable regions. *Nature* **312**:40-46.

16. Barth, R., B. Kim, N. Lan, T. Hunkapiller, N. Sobieck, A. Winoto, H. Gershenfeld, C. Okada, D. Hansburg, I. Weissman, and L. Hood. 1985. The murine T-cell receptor employs a limited repertoire of expressed V_β gene segments. *Nature* **316**:517-523.

17. Behlke, M. A., D. G. Spinella, H. S. Chou, W. Sha, D. L. Hartl, and D. Y. Loh. 1985. T-cell receptor β chain expression: Dependence on relatively few variable region genes. *Science* **229**:566-570.

18. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.

19. Maniatis, T., E. F. Fritsch, and J. Sambrook. J. *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

20. Rigby, P.W.J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* **113**:237-251.

21. Kobori, J. A., A. Winoto, J. McNicholas, and L. Hood. 1984. Molecular characterization of the recombination region of six murine major hostocompatibility complex (MHC) I-region recombinants. *J. Mol. Cell Immunol.* **1**:125-131.

22. Smith, H. O. and M. L. Birnstiel. 1976. A simple method for DNA restriction site mapping. *Nucl. Acids Res.* **3**:2387-2398.

23. Maxam, A. M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Meth. Enzymol.* **65**:499-560.

24. Strauss, E., J. Kobori, G. Siu, and L. Hood. 1985. Specific-primer-directed DNA sequencing. *Analyt. Biochem.* (submitted).

25. Acuto, O., M. Fabbi, J. Smart, C. B. Poole, J. Protentis, H. D. Royer, S. F. Schlossman, and E. L. Reinherz. 1984. Purification and N-terminal amino acid sequencing of the β subunit of a human T-cell antigen receptor. *Proc. Natl. Acad. Sci. USA* **81**:3851-3855.

26. Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci. USA* **78**:454-458.

27. Miyata, T., T. Yasunaga, and T. Nishida. 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. *Proc. Natl. Acad. Sci. USA* **77**:7328-7332.

28. Tunnacliffe, A., R. Kefford, C. Milstein, A. Forster, and T. H. Rabbitts. 1985. Sequence and evolution of the human T-cell receptor β-chain genes. *Proc. Natl. Acad. Sci. USA* **82**:5068-5072.

29. Yoshikai, Y., D. Anatoniou, S. P. Clark, Y. Yanagi, R. Sangster, P. Van den Elsen, C. Terhorst, and T. W. Mak. 1984. Sequence and expression of transcripts of the human T-cell receptor β-chain genes. *Nature* **312**: 521-524.

30. Jones, N., J. Leiden, D. Dialynas, J. Fraser, M. Clabby, T. Kishimoto, J. L. Strominger, D. Andrews, W. Lane, and J. Woody. 1984. Partial primary structure of the alpha and beta chains of human tumor T-cell receptors. *Science* **227**:311-314.

31. Acuto, O., T. J. Campen, H. D. Royer, R. E. Hussey, C. B. Poole, and E. L. Reinherz. 1985. Molecular analysis of T-cell receptor (Ti) variable region (V) gene expression. Evidence that a single Ti β V gene family can be used in formation of V domains on phenotypically and functionally diverse T-cell populations. *J. Exp. Med.* **161**:1326-1343.

32. Sims, J. E., A. Tunnacliffe, W. J. Smith, and T. H. Rabbitts. 1984. Complexity of human T-cell antigen receptor β-chain constant- and variable-region genes. *Nature* **312**:541-545.

33. Goverman, J., K. Minard, N. Shastri, T. Hunkapiller, D. Hansburg, E. Sercarz, and L. Hood. 1985. Rearranged β T-cell receptor genes in a helper T-cell clone specific for lysozyme: No correlation between V_β and MHC restriction. *Cell* **40**:859-867.

34. Segal, D., E. Padlan, G. Cohen, S. Rudikoff, M. Potter, and D. Davies. 1974. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Natl. Acad. Sci. USA* **71**:4298-4302.

35. Saul, F., L. Amzel, and R. Poljak. 1978. Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin new at 2.0 Å resolution. *J. Biol. Chem.* **253**:585-597.

36. Amzel, L. M. and R. J. Poljak. 1979. Three-dimensional structure of immunoglobulins. *Ann. Rev. Biochem.* **48**:961-997.

37. Hood, L., J. H. Campbell, and S.C.R. Elgin. (1965). The organization, expression, and evolution of antibody genes and other multigene families. *Ann. Rev. Genetics* **9**:305-353.

38. Breathnach, R. and P. Chambon. 1981. Organization and expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* **50**:349-383.

39. Hayday, A. C., H. Saito, S. D. Gillies, D. M. Kranz, G. Tanigawa, H. N. Eisen, and S. Tonegawa. 1985. Structure, organization and somatic rearrangement of T-cell gamma genes. *Cell* **40**:259-269.

40. Malissen, M., C. McCoy, D. Blanc., J. Trucy, C. Devaux, A.-M. Schmitt-Verhulst, F. Fitch, L. Hood, and B. Malissen. 1985. Direct evidence for chromosomal inversion during T-cell receptor β gene rearrangements. *Nature*, in press.

41. Falkner, F. G. and H. G. Zachau. 1984. Correct transcription of an immunoglobulin κ gene requires an upstream fragment containing conserved sequence elements. *Nature* **310**:71-74.

42. Parslow, T. G., D. L. Blair, W. J. Murphy, and D. K. Granner. 1984. Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc. Natl. Acad. Sci. USA* **81**:2650-2654.

43. Grosschedl, R. and D. Baltimore. 1985. Cell-type specificity of immunoglobulin gene expression is regulated by at least three DNA sequence elements. *Cell* **41**:885-897.

44. Baltimore, D. 1981. Gene conversion: Some implications for immunoglobulin genes. *Cell* **24**:592-594.

45. Salser, W. 1976. Globin mRNA sequences: Analysis of Base Pairing and evolutionary implications. *Cold Spring Harbor Symp. Quant. Biol.* **62**:985-1002.

46. Miyata, T., H. Hayashida, R. Kikuno, M. Hasegawa, M. Kobayashi, and K. Koike. 1982. Molecular clock of silent substitution: at least six-fold preponderence of silent changes in mitochondrial genes over those in nuclear genes. *J. Mol. Evol.* **19**:28-35.

47. Miyata, T. and H. Hayashida. 1982. Recent divergence from a common ancestor of human IFN-α genes. *Nature* **295**:165-168.

48. Li, W.-H., C.-I. Wu, and C.-C. Luo. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**:150-174.

49. Federoff, N. V. 1979. On spacers. *Cell* **16**:697-710.

Fig. 1. Genomic blots on human germline DNA with the YT35 V-region-specific probe. The DNA samples were digested with the restriction enzymes Eco RI or Bam HI. The sizes of the bands are indicated, along with the corresponding V_β gene segment.

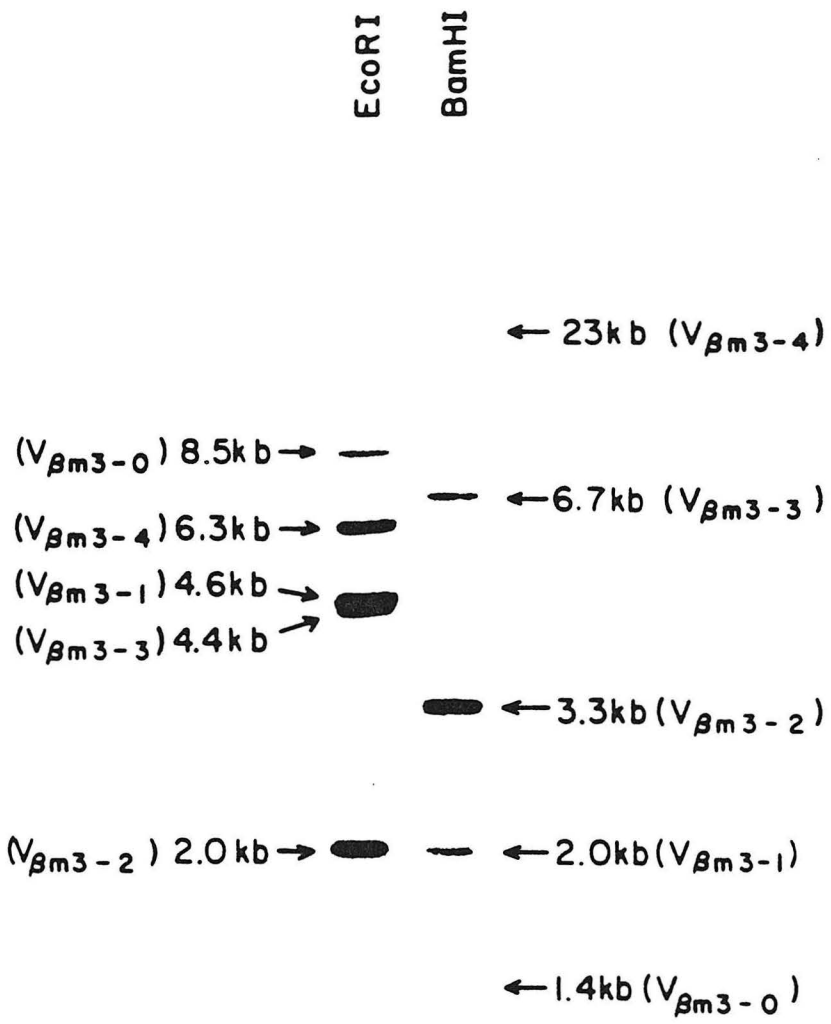
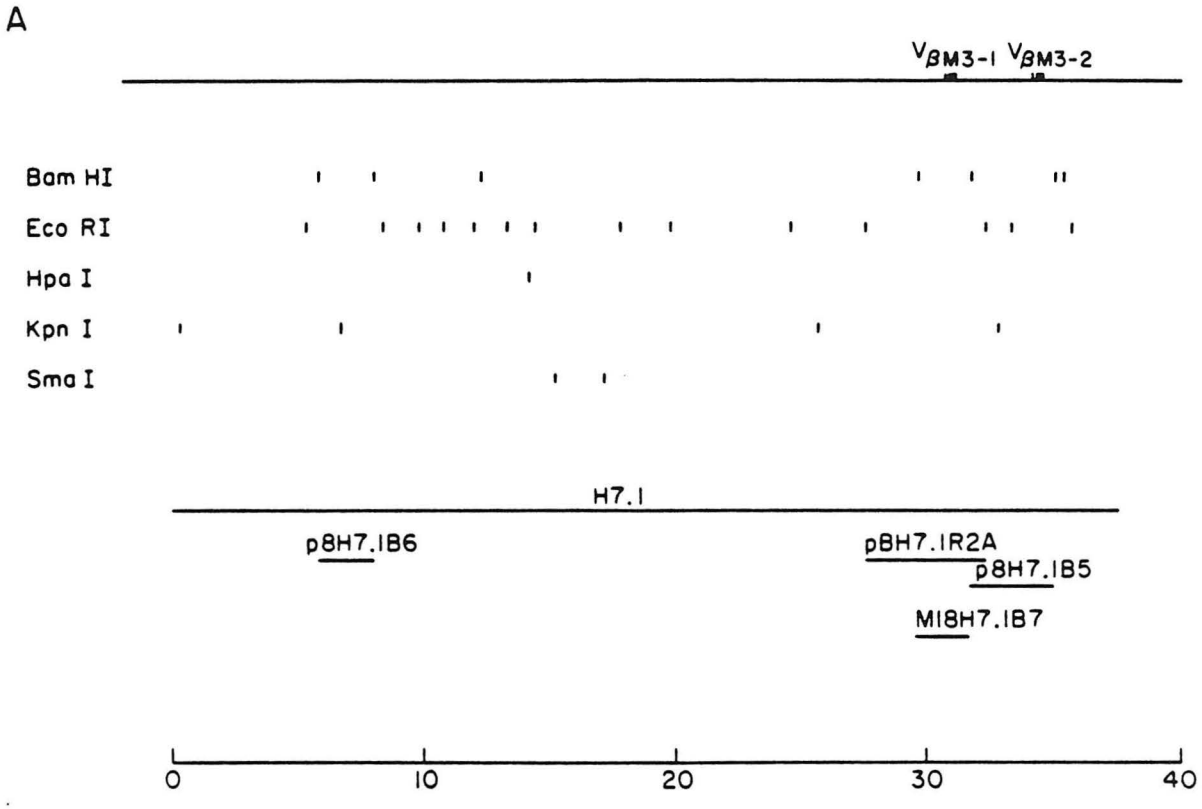
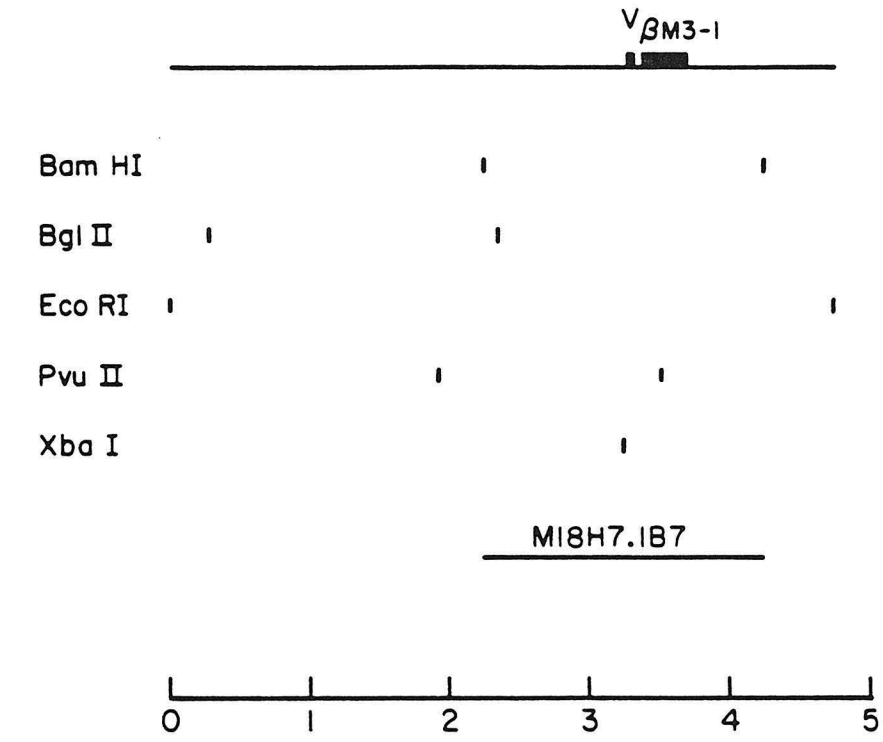

EcoRI
BamHI
← 23kb ($V_{\beta m3-4}$)
($V_{\beta m3-0}$) 8.5kb →
($V_{\beta m3-4}$) 6.3kb →
($V_{\beta m3-1}$) 4.6kb →
($V_{\beta m3-3}$) 4.4kb →
← 6.7kb ($V_{\beta m3-3}$)
← 3.3kb ($V_{\beta m3-2}$)
($V_{\beta m3-2}$) 2.0kb →
← 2.0kb ($V_{\beta m3-1}$)
← 1.4kb ($V_{\beta m3-0}$)

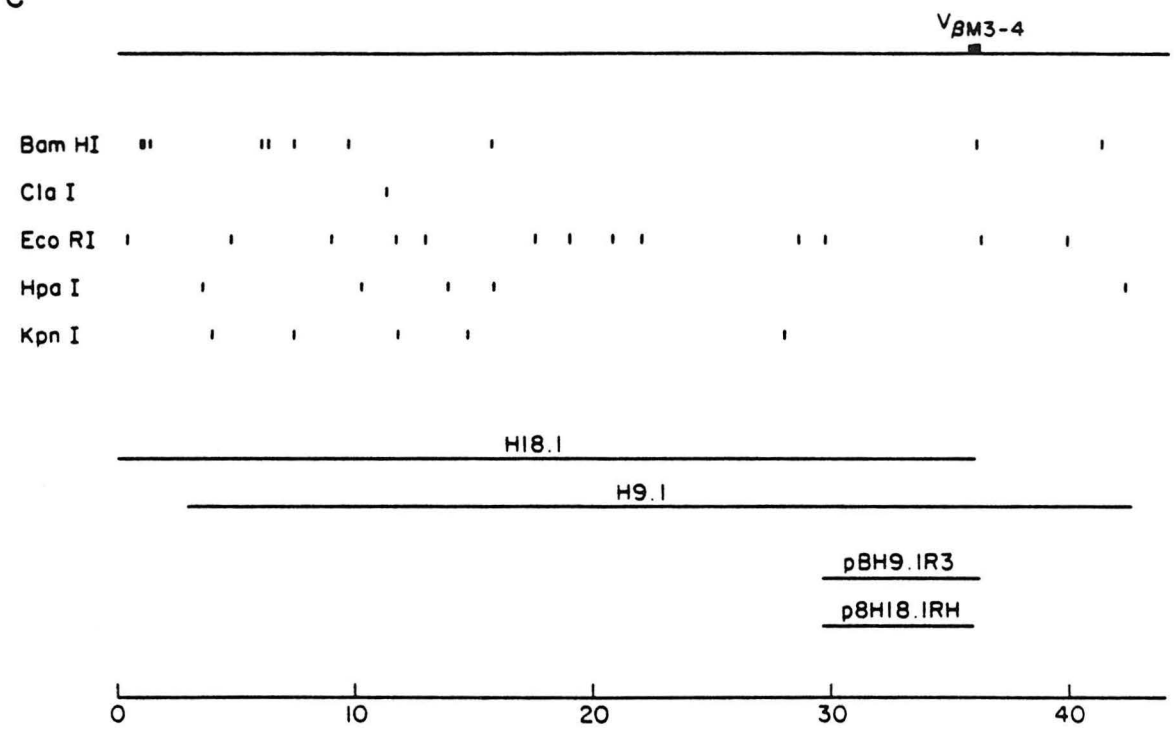
Fig. 2. Restriction maps of the cosmid and lambda clones containing the members of the $V_{\beta M3}$ family. A) Cosmid H7.1, containing $V_{\beta M3-1}$ and $V_{\beta M3.2}$, and the pBR322 subclone pBH7.1R2A, the pUC8 subclones p8H7.1B6 and p8H7.1B5, and the M13mp18 subclone M18H7.1B7 of this region, B) Plasmid pBH7.1R2A, containing the $V_{\beta M3-1}$ gene segment, and the M13mp18 subclone M18H7.1B7, C) Cosmids H18.1 and H9.1, containing $V_{\beta M3-4}$, and the pBR322 subclone pBH9.1R3 and the pUC8 subclone p8H18.1RH that contains portions of these cosmids, D) λVBM3.0, containing the $V_{\beta M3-0}$ gene segment, and the M13mp18 subclone M18VBM30 that contains a portion of this region, E) λgt clone λVBM3.3, containing the $V_{\beta M3-3}$ gene segment. Scale in each figure is in kilobases. The thin and thick boxes represent the first and second exons of the germline V_{β} gene segment.


A
$V_{\beta M3-1}$ $V_{\beta M3-2}$
Bam HI
Eco RI
Hpa I
Kpn I
Sma I
H7.1
p8H7.1B6
pBH7.1R2A
p8H7.1B5
M18H7.1B7
0
10
20
30
40

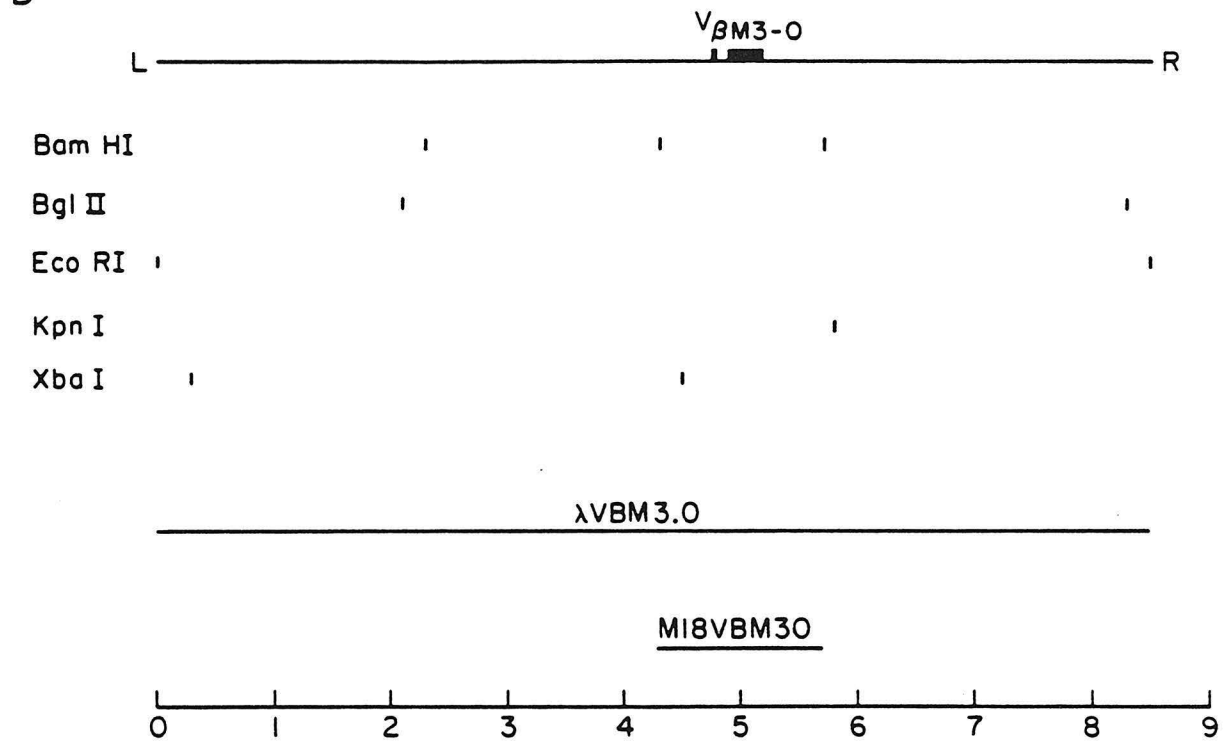
B



C


Vβm3-4
Bam HI
Cla I
Eco RI
Hpa I
Kpn I
H18.1
H9.1
pBH9.1R3
p8H18.1RH
0
10
20
30
40

D


$V_{\beta M3-0}$
L
R
Bam HI
Bgl II
Eco RI
Kpn I
Xba I
λVBM3.0
MI8VBM30
0
1
2
3
4
5
6
7
8
9

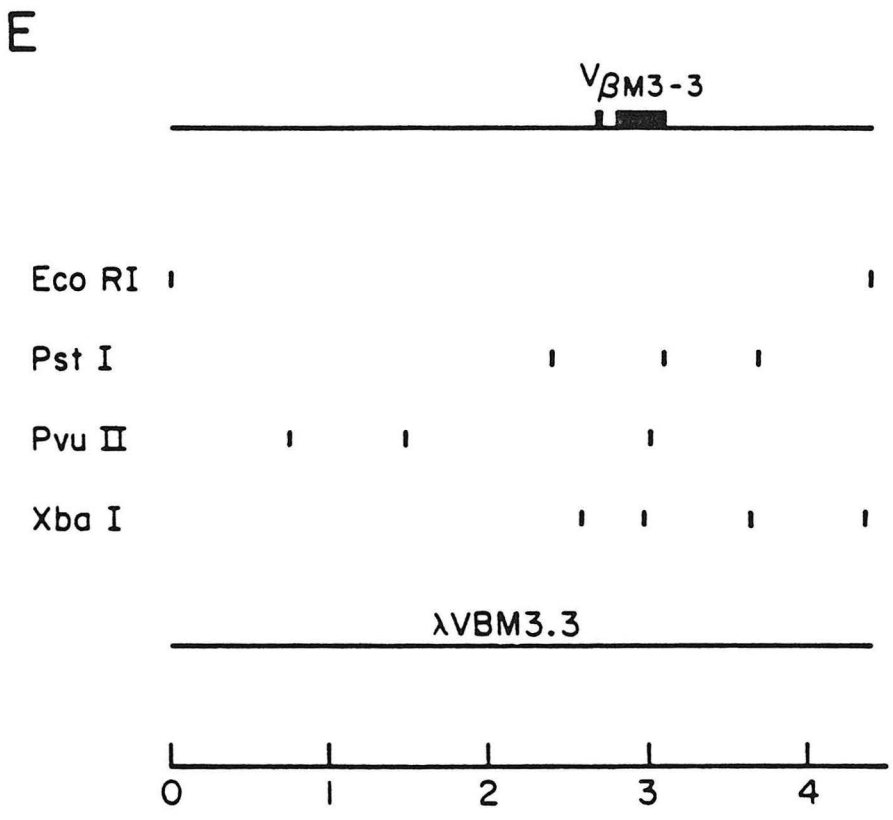

E
V_{β}M3-3
Eco RI
Pst I
Pvu II
Xba I
λVBM3.3
0
1
2
3
4

Fig. 3. Complete nucleotide sequence of the five members of the $V_{\beta M3}$ family and their flanking regions. Gaps were introduced to maximize homology. Leader and V coding regions are boxed and indicated. Putative TATA and CCAAT boxes in the 5' region and the rearrangement recognition signals in the 3' flanking region are overlined and indicated. The 21 bp direct repeat in the 5' flanking region of $V_{\beta M3\text{-}0}$ is indicated by two arrows, and the 16 bp conserved region in the 5' flanking region of the $V_{\beta M3}$ subfamily members is indicated by a bracket.

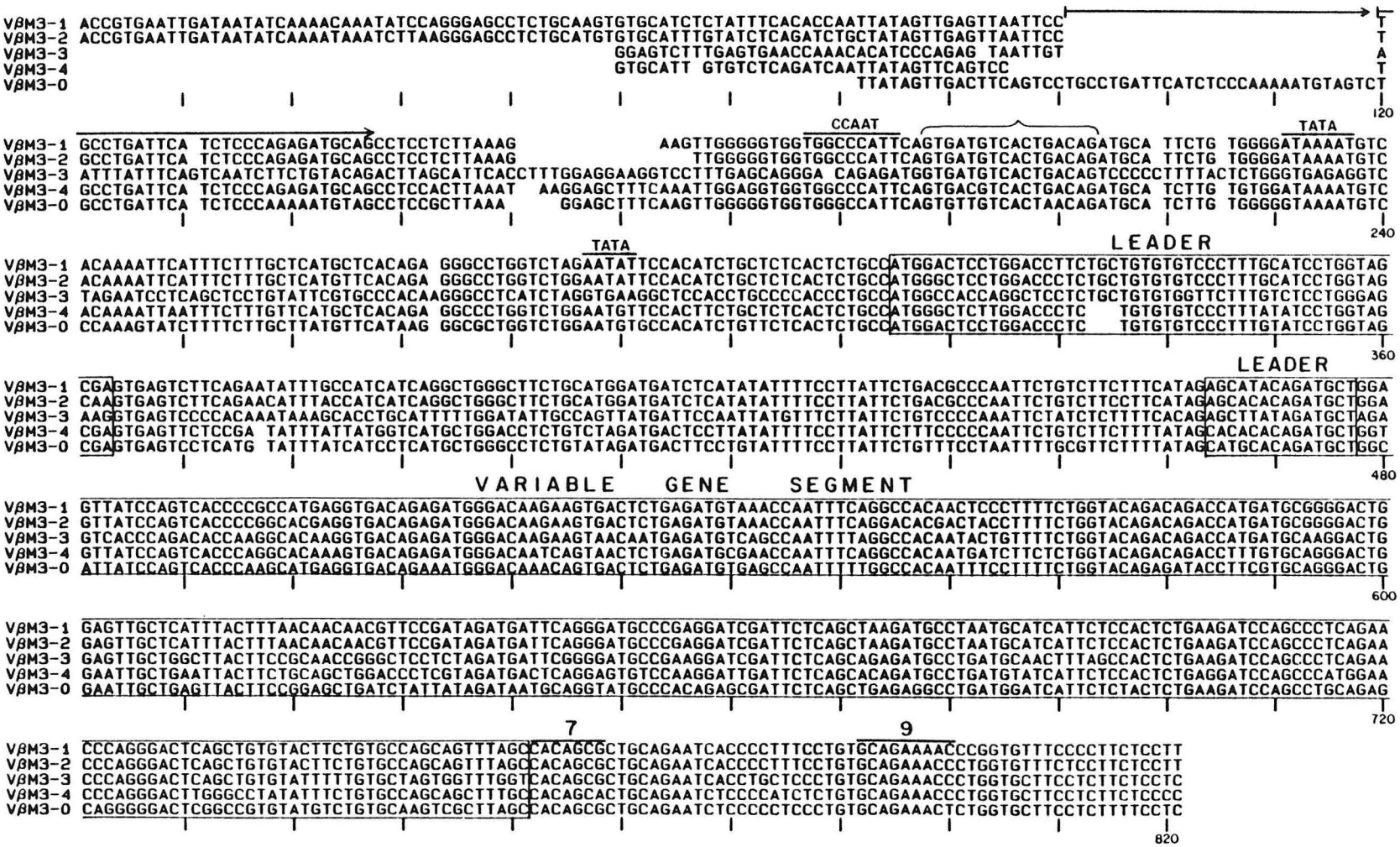

VβM3-1 ACCGTGAATTGATAATATCAAAACAAATATCCAGGGAGCCTCTGCAAGTGTGCATCTCTATTTCACACCAATTATAGTTGAGTTAATTCC T
VβM3-2 ACCGTGAATTGATAATATCAAAATAAATCTTAAGGGAGCCTCTGCATGTGTGCATTTGTATCTCAGATCTGCTATAGTTGAGTTAATTCC T
VβM3-3 GGAGTCTTTGAGTGAACCAAACACATCCCAGAG TAATTGT A
VβM3-4 GTGCATT GTGTCTCAGATCAATTATAGTTCAGTCC T
VβM3-0 TTATAGTTGACTTCAGTCCTGCCTGATTCATCTCCCAAAAATGTAGTCT
120
CCAAT
TATA
VβM3-1 GCCTGATTCA TCTCCCAGAGATGCAGCCTCCTCTTAAAG AAGTTGGGGGTGGTGGCCCATTCAGTGATGTCACTGACAGATGCA TTCTG TGGGGATAAAATGTC
VβM3-2 GCCTGATTCA TCTCCCAGAGATGCAGCCTCCTCTTAAAG TTGGGGGTGGTGGCCCATTCAGTGATGTCACTGACAGATGCA TTCTG TGGGGATAAAATGTC
VβM3-3 ATTTATTTCAGTCAATCTTCTGTACAGACTTAGCATTCACCTTTGGAGGAAGGTCCTTTGAGCAGGGA CAGAGATGGTGATGTCACTGACAGTCCCCCTTTTACTCTGGGTGAGAGGTC
VβM3-4 GCCTGATTCA TCTCCCAGAGATGCAGCCTCCACTTAAAT AAGGAGCTTTCAAATTGGAGGTGGTGGCCCATTCAGTGACGTCACTGACAGATGCA TCTTG TGTGGATAAAATGTC
VβM3-0 GCCTGATTCA TCTCCCAAAAATGTAGCCTCCGCTTAAA GGAGCTTTCAAGTTGGGGGTGGTGGGCCATTCAGTGTTGTCACTAACAGATGCA TCTTG TGGGGGTAAAATGTC
240
TATA
LEADER
VβM3-1 ACAAAATTCATTTCTTTGCTCATGCTCACAGA GGGCCTGGTCTAGAATATTCCACATCTGCTCTCACTCTGCCATGGACTCCTGGACCTTCTGCTGTGTGTCCCTTTGCATCCTGGTAG
VβM3-2 ACAAAATTCATTTCTTTGCTCATGTTCACAGA GGGCCTGGTCTGGAATATTCCACATCTGCTCTCACTCTGCCATGGGCTCCTGGACCCTCTGCTGTGTGTCCCTTTGCATCCTGGTAG
VβM3-3 TAGAATCCTCAGCTCCTGTATTCGTGCCCACAAGGGCCTCATCTAGGTGAAGGCTCCACCTGCCCCACCCTGCCATGGCCACCAGGCTCCTCTGCTGTGTGGTTCTTTGTCTCCTGGGAG
VβM3-4 ACAAAATTAATTTCTTTGTTCATGCTCACAGA GGCCCTGGTCTGGAATGTTCCACTTCTGCTCTCACTCTGCCATGGGCTCTTGGACCCTC TGTGTGTCCCTTTATATCCTGGTAG
VβM3-0 CCAAAGTATCTTTTCTTGCTTATGTTCATAAG GGCGCTGGTCTGGAATGTGCCACATCTGTTCTCACTCTGCCATGGACTCCTGGACCCTC TGTGTGTCCCTTTGTATCCTGGTAG
360
LEADER
VβM3-1 CGAGTGAGTCTTCAGAATATTTGCCATCATCAGGCTGGGCTTCTGCATGGATGATCTCATATATTTTCCTTATTCTGACGCCCAATTCTGTCTTCTTTCATAGAGCATACAGATGCTGGA
VβM3-2 CAAGTGAGTCTTCAGAACATTTACCATCATCAGGCTGGGCTTCTGCATGGATGATCTCATATATTTTCCTTATTCTGACGCCCAATTCTGTCTTCCTTCATAGAGCACACAGATGCTGGA
VβM3-3 AAGGTGAGTCCCCACAAATAAAGCACCTGCATTTTTGGATATTGCCAGTTATGATTCCAATTATGTTTCTTATTCTGTCCCCAAATTCTATCTCTTTTCACAGAGCTTATAGATGCTAGA
VβM3-4 CGAGTGAGTTCTCCGA TATTTATTATGGTCATGCTGGACCTCTGTCTAGATGACTCCTTATATTTTCCTTATTCTTTCCCCCAATTCTGTCTTCTTTTATAGCCACACAGATGCTGGT
VβM3-0 CGAGTGAGTCCTCATG TATTTATCATCCTCATGCTGGGCCTCTGTATAGATGACTTCCTGTATTTTCCTTATTCTGTTTCCTAATTTTGCGTTCTTTTATAGCATGCACAGATGCTGGC
480
VARIABLE GENE SEGMENT
VβM3-1 GTTATCCAGTCACCCCGCCATGAGGTGACAGAGATGGGACAAGAAGTGACTCTGAGATGTAAACCAATTTCAGGCCACAACTCCCTTTTCTGGTACAGACAGACCATGATGCGGGGACTG
VβM3-2 GTTATCCAGTCACCCCGGCACGAGGTGACAGAGATGGGACAAGAAGTGACTCTGAGATGTAAACCAATTTCAGGACACGACTACCTTTTCTGGTACAGACAGACCATGATGCGGGGACTG
VβM3-3 GTCACCCAGACACCAAGGCACAAGGTGACAGAGATGGGACAAGAAGTAACAATGAGATGTCAGCCAATTTTAGGCCACAATACTGTTTTCTGGTACAGACAGACCATGATGCAAGGACTG
VβM3-4 GTTATCCAGTCACCCAGGCACAAAGTGACAGAGATGGGACAATCAGTAACTCTGAGATGCGAACCAATTTCAGGCCACAATGATCTTCTCTGGTACAGACAGACCTTTGTGCAGGGACTG
VβM3-0 ATTATCCAGTCACCCAAGCATGAGGTGACAGAAATGGGACAAACAGTGACTCTGAGATGTGAGCCAATTTTTGGCCACAATTTCCTTTTCTGGTACAGAGATACCTTCGTGCAGGGACTG
600
VβM3-1 GAGTTGCTCATTTACTTTAACAACAACGTTCCGATAGATGATTCAGGGATGCCCGAGGATCGATTCTCAGCTAAGATGCCTAATGCATCATTCTCCACTCTGAAGATCCAGCCCTCAGAA
VβM3-2 GAGTTGCTCATTTACTTTAACAACAACGTTCCGATAGATGATTCAGGGATGCCCGAGGATCGATTCTCAGCTAAGATGCCTAATGCATCATTCTCCACTCTGAAGATCCAGCCCTCAGAA
VβM3-3 GAGTTGCTGGCTTACTTCCGCAACCGGGCTCCTCTAGATGATTCGGGGATGCCGAAGGATCGATTCTCAGCAGAGATGCCTGATGCAACTTTAGCCACTCTGAAGATCCAGCCCTCAGAA
VβM3-4 GAATTGCTGAATTACTTCTGCAGCTGGACCCTCGTAGATGACTCAGGAGTGTCCAAGGATTGATTCTCAGCACAGATGCCTGATGTATCATTCTCCACTCTGAGGATCCAGCCCATGGAA
VβM3-0 GAATTGCTGAGTTACTTCCGGAGCTGATCTATTATAGATAATGCAGGTATGCCCACAGAGCGATTCTCAGCTGAGAGGCCTGATGGATCATTCTCTACTCTGAAGATCCAGCCTGCAGAG
720
7
9
VβM3-1 CCCAGGGACTCAGCTGTGTACTTCTGTGCCAGCAGTTTAGCCACAGCGCTGCAGAATCACCCCTTTCCTGTGCAGAAAACCCGGTGTTTCCCCTTCTCCTT
VβM3-2 CCCAGGGACTCAGCTGTGTACTTCTGTGCCAGCAGTTTAGCCACAGCGCTGCAGAATCACCCCTTTCCTGTGCAGAAACCCTGGTGTTTCTCCTTCTCCTT
VβM3-3 CCCAGGGACTCAGCTGTGTATTTTTGTGCTAGTGGTTTGGTCACAGCGCTGCAGAATCACCTGCTCCCTGTGCAGAAACCCTGGTGCTTCCTCTTCTCCTC
VβM3-4 CCCAGGGACTTGGGCCTATATTTCTGTGCCAGCAGCTTTGCCACAGCACTGCAGAATCTCCCCATCTCTGTGCAGAAACCCTGGTGCTTCCTCTTCTCCCC
VβM3-0 CAGGGGGACTCGGCCGTGTATGTCTGTGCAAGTCGCTTAGCCACAGCGCTGCAGAATCTCCCCCTCCCTGTGCAGAAACTCTGGTGCTTCCTCTTTTCCTC
820

Fig. 4. Protein alignments of the $V_{\beta M3}$ family compared with all of the human V_β genes published to date and the murine $V_{\beta 11}$ gene segment. Gaps were aligned to maximize sequence similarity. Conserved positions in human and mouse V_β gene segments (>75% frequency of a single amino acid) are indicated by an asterisk (see Kronenberg et al., 1986). Amino acids conserved in sequence and position with other V gene families are indicated by letters in an open box: L = immunoglobulin κ and λ light chain V gene family, H = immunoglobulin heavy chain V gene family, and α = V_α gene family. Invariant amino acids thought to interact with one another to stabilize V_H-V_L interactions in immunoglobulins are indicated with an arrow. The references for the previously published V_β gene sequences are: $V_{\beta 11}$ (17), HPB-MLT (30), HPB-ALL (29), and MOLT-4 (28).

Vβ M3-1 GVIQSPRHEVTEMGQEVTLRCKPISGH NSLFWYRQTMMRGLELLIYFNNNV PIDDSGMPEDRFSAKMPNASFSTLKIQPSEPRDSAVYFCASSL
Vβ M3-2 GVIQSPRHEVTEMGQEVTLRCKPISGH DYLFWYRQTMMRGLELLIYFNNNV PIDDSGMPEDRFSAKMPNASFSTLKIQPSEPRDSAVYFCASSL
Vβ M3-3 RVTQTPRHKVTEMGQEVTMRCQPILGH NSVFWYRQTMMQGLELLAYFANRA PLDDSGMPKDRFSAEMPDATLATLKIQPSEPRDSAVYFCASGL
Vβ M3-4 GVIQSPRHKVTEMGQSVTLRCEPISGH NDLLWYRQTFVQGLELLNYFCSWT LVDDSGVSKDXFSAQMPDVSFSTLRIQPMEPRDLGLYFCASSF
Vβ M3-0 GIIQSPKHEVTEMGQTVTLRCEPILAT ISFFWYRDTFVQGLELLSYFRSXS IIDNAGMPTERFSAERPDGSFSTLKIQPAEQGDSAVYVCASRL
Vβ 11 GVIQTPRHKVTGKGQEATLWCEPISGH SAVFWYRQTIVQGLEFLTYFRNQA PIDDSGMPKERFSAQMPNQSHSTLKIQSTQPQDSAVYLCASS
HPB-MLT WELRCDPISGH TALYWYRQSLGQGPELLIYFQGTG AADDSGLPNDRFFAVRPEGSVSTLKIQRTERGDSAVYLCASS
HPB-ALL EEESFPDRFSARQFPNYSSELNVNALLLGDSALYLCASS
MOLT4 VVSQHPSWVICKSGTSVKIECRSLDFQATTMFWYRQFPKQSLMLMATSNEGSKATYEQGVEKDKFLINHASLTLSTLTVTSAHPEDSSFYICSAR
10
20
30
40
50
60
70
80
90

TABLE I

Comparison	5' Flanking	Leader	Intron	V Gene Segment	3' Flanking
$V_{\beta M3-1}$ vs. $V_{\beta M3-2}$	94.3%	93.9%	97.0%	97.7%	95.2%
$V_{\beta M3-1}$ vs. $V_{\beta M3-3}$	46.6%	69.4%	59.0%	76.6%	83.9%
$V_{\beta M3-1}$ vs. $V_{\beta M3-4}$	89.2%	89.2%	77.8%	78.8%	79.4%
$V_{\beta M3-1}$ vs. $V_{\beta M3-0}$	84.2%	95.7%	77.8%	77.5%	81.0%
$V_{\beta M3-2}$ vs. $V_{\beta M3-3}$	46.3%	73.5%	57.0%	75.6%	76.4%
$V_{\beta M3-2}$ vs. $V_{\beta M3-4}$	90.5%	91.3%	76.8%	79.1%	80.2%
$V_{\beta M3-2}$ vs. $V_{\beta M3-0}$	84.5%	93.5%	76.8%	77.0%	80.5%
$V_{\beta M3-3}$ vs. $V_{\beta M3-4}$	42.8%	67.4%	56.6%	71.7%	81.2%
$V_{\beta M3-3}$ vs. $V_{\beta M3-0}$	46.8%	71.7%	51.5%	68.7%	83.2%
$V_{\beta M3-4}$ vs. $V_{\beta M3-0}$	84.0%	93.5%	80.8%	79.1%	89.5%

Comparison of the coding and flanking regions of the $V_{\beta M3}$ subfamily members. 5' flanking region is defined as bases 1-314 in Figure 3, the leader as bases 315-363 and bases 464-477, the intron as bases 364-463, the V gene segment as bases 478-761, and the 3' flanking region as bases 762-894.

TABLE II

Mutation Frequencies in the $V_{\beta M3}$ Family

Comparison	K_1*	K_2*	K_3*	K_A+	K_S+	K_A/K_S
$V_{\beta M3-1}$ vs $V_{\beta M3-2}$	0.011 ± 0.01	0.011 ± 0.01	0.038 ± 0.02	0.0092 ± 0.006	0.049 ± 0.03	0.19
$V_{\beta M3-1}$ vs $V_{\beta M3-3}$	0.22 ± 0.05	0.092 ± 0.03	0.29 ± 0.06	0.14 ± 0.03	0.47 ± 0.12	0.29
$V_{\beta M3-1}$ vs $V_{\beta M3-4}$	0.25 ± 0.06	0.17 ± 0.05	0.33 ± 0.08	0.19 ± 0.03	0.50 ± 0.14	0.38
$V_{\beta M3-1}$ vs $V_{\beta M3-0}$	0.30 ± 0.07	0.20 ± 0.05	0.39 ± 0.08	0.22 ± 0.04	0.55 ± 0.15	0.40
$V_{\beta M3-2}$ vs $V_{\beta M3-3}$	0.24 ± 0.06	0.092 ± 0.03	0.30 ± 0.07	0.14 ± 0.03	0.50 ± 0.13	0.28
$V_{\beta M3-2}$ vs $V_{\beta M3-4}$	0.28 ± 0.07	0.14 ± 0.04	0.33 ± 0.08	0.19 ± 0.03	0.52 ± 0.15	0.37
$V_{\beta M3-2}$ vs $V_{\beta M3-0}$	0.30 ± 0.07	0.20 ± 0.05	0.39 ± 0.08	0.23 ± 0.04	0.56 ± 0.14	0.41
$V_{\beta M3-3}$ vs $V_{\beta M3-4}$	0.33 ± 0.07	0.16 ± 0.05	0.36 ± 0.08	0.22 ± 0.04	0.57 ± 0.17	0.39
$V_{\beta M3-3}$ vs $V_{\beta M3-0}$	0.36 ± 0.08	0.19 ± 0.05	0.49 ± 0.10	0.25 ± 0.04	0.77 ± 0.23	0.33
$V_{\beta M3-4}$ vs $V_{\beta M3-0}$	0.32 ± 0.07	0.18 ± 0.05	0.45 ± 0.10	0.23 ± 0.04	0.71 ± 0.22	0.33
$V_{\beta 11}$ vs $V_{\beta M3-2}$	0.25 ± 0.06	0.16 ± 0.05	0.43 ± 0.09	0.18 ± 0.03	0.75 ± 0.24	0.24

*K_1, K_2 and K_3 represent the mutation frequencies in the first, second and third positions of the codon, respectively.

+K_A and K_S represent the mutation frequencies in the amino-acid replacement- and silent-sites.

Error limits are one standard deviation.

TABLE III

*Homology Matrix of V_β Gene Segments**

	$V_{\beta M3-1}$	$V_{\beta M3-2}$	$V_{\beta 3-3}$	$V_{\beta M3-4}$	$V_{\beta M3-0}$	Murine $V_{\beta 11}$	HPB-MTL	HPB-ALL	MOLT4
$V_{\beta M3-1}$	-	97.9	77.7	70.2	66.0	69.9	61.0	38.5	26.9
$V_{\beta M3-2}$	98.2	-	75.5	69.2	64.9	69.9	61.0	38.5	26.9
$V_{\beta M3-3}$	82.1	82.1	-	63.8	60.6	66.7	53.3	33.3	30.1
$V_{\beta M3-4}$	79.2	79.6	77.1	-	59.6	62.4	53.3	33.3	28.0
$V_{\beta M3-0}$	76.4	75.4	72.9	76.1	-	59.1	53.3	35.9	25.8
Murine $V_{\beta 11}$	77.4	77.8	77.4	74.9	71.3	-	58.5	38.5	25.8
HPB-MLT	70.6	69.7	65.4	67.1	68,0	68.4	-	41.0	24.7
HPB-ALL	39.1	38.0	39.8	38.7	35.9	38.7	41.1	-	25.7
MOLT4	54.2	54.2	49.2	55.9	56.8	54.7	54.7	32.5	-

*Numbers above the diagonal designate the percentage homology of sequences on the X and Y axes when compared at the protein level; numbers below the diagonal show percentage homology at the DNA level.

CHAPTER FIVE

THE STRUCTURE, REARRANGEMENT, AND EXPRESSION OF D_β GENE SEGMENTS OF THE MURINE T-CELL ANTIGEN RECEPTOR

Published in *Nature*

Reprinted from Nature, Vol. 311, No. 5984, pp. 344-350, 27 September 1984

The structure, rearrangement and expression of D_β gene segments of the murine T-cell antigen receptor

Gerald Siu*, Mitchell Kronenberg*, Erich Strauss*, Regina Haars†, Tak W. Mak‡ & Leroy Hood*

* Division of Biology, California Institute of Technology, Pasadena, California 91125, USA
† Department of Pathology, College of Physicians and Surgeons of Columbia University, New York 10032, USA
‡ Department of Medical Biophysics, University of Toronto, Ontario Cancer Institute, Toronto, Canada M4X 1K9

It has been postulated that the variable region of the β-polypeptide of the murine T-cell antigen receptor is encoded by three distinct germ-line gene segments—variable (V_β), diversity (D_β) and joining (J_β)—that are rearranged to generate a V_β gene. Germ-line V_β and J_β gene segments have been isolated previously. Here we report the isolation and characterization of two germ-line D_β gene segments that have recognition signals for DNA rearrangement strikingly similar to those found in the three immunoglobulin gene families and in V_β and J_β gene segments. The D_β and J_β segments can join in the absence of V_β gene segment rearrangement and these rearranged sequences are transcribed in some T cells.

IMMUNOGLOBULINS and T-cell antigen receptor molecules exhibit extensive diversity but the genetic basis and the mechanisms responsible for generating this diversity are well understood only for the immunoglobulin genes[1-3]. Immunoglobulins are heterodimers made up of disulphide-bridged light and heavy chains each of which is divided into variable regions that recognize antigen, and constant regions that control effector functions. Genes that encode the variable region of light chains are created by a DNA rearrangement that joins two distinct gene segments—variable (V_L) and joining (J_L)[4-6]. The variable region genes of heavy chains are created by the joining of three distinct gene segments—variable (V_H), diversity (D_H), and joining (J_H)[7,8]. One mechanism for the generation of immunoglobulin diversity is joining of different V_H, D_H, and J_H, or V_L and J_L gene segments[9,10]. Also gene segments may be joined together at different sites, leading to a junctional variability[10-12] and in the heavy chain gene family, extra nucleotides, not found in V_H, D_H, or J_H gene segments, may be added between these gene segments as a result of the joining process (N-region diversity)[11,13]. The 3′ side of the V gene segments, the 5′ side of the J gene segments, and both sides of the D gene segments are flanked by recognition signals for DNA rearrangements. These signals comprise three components—a highly conserved palindromic heptamer, a nonconserved spacer sequence of 12 or 23 nucleotides, and a relatively conserved A/T-rich nonamer[6-8,14]. It is believed that these recognition sequences play a critical role in DNA rearrangements because they are present in all three of the immunoglobulin gene families and, therefore, have been conserved throughout vertebrate evolution.

The T-cell antigen receptor is a heterodimer composed of two disulphide-bonded polypeptides, α and β, that have a similar molecular weight[15-17]. The genes that encode the β-chain of the T-cell antigen receptor have recently been identified and they share many features with those encoding immunoglobulins. The nucleotide sequence of cDNA clones encoding the β-chain of the T-cell receptor from both mouse and man contains regions similar to V and J gene segments and C genes[18-20]. Distinct germ-line V_β and J_β gene segments and C_β genes have been isolated from genomic libraries[21-23]. Two highly similar and closely-linked C_β genes denoted $C_{\beta1}$ and $C_{\beta2}$ have been isolated on a single cosmid clone of germ-line DNA and each C_β gene has a cluster of J_β gene segments associated with it[23]. The V_β and J_β gene segments have recognition signals for DNA rearrangements that are very similar to those of their immunoglobulin counterparts[21-23].

Both murine and human β cDNA clones encoding a complete variable region contain eight or more nucleotides than their corresponding germ-line V_β and J_β gene segments suggesting that a portion of the β variable region is encoded by a third gene segment analogous to the D_H gene segment of immunoglobulins[21,22]. Indeed, in this paper we present the structure of two germline D_β gene segments. One of these D_β gene segments lies to the 5′ side of the $J_{\beta1}$ gene cluster and the second lies to the 5′ side of the $J_{\beta2}$ gene cluster, between $C_{\beta1}$ and $J_{\beta2}$. Recognition signals for DNA rearrangements on either side of these D gene segments are described, as are rearrangements and transcripts involving the D_β gene segments.

D_β gene segment of each J_β gene cluster

The partial nucleotide sequences of four murine β-chain cDNA clones have been published[19]. Two of them appeared to lack a V_β gene segment but contain joined D_β–J_β sequences. This conclusion was based on several features. First, the published sequences 5′ to the J_β gene segments resemble neither V_β gene segments nor the sequences found 5′ to the germ-line J_β gene segments. Second, in both of these clones the sequences 5′ to the joined J_β gene segment have a recognition signal for DNA joining that includes the heptamer, a 12-nucleotide spacer sequence, and the A/T-rich nonamer sequence, as well as a short sequence that could be a D_β coding segment located between the joining signal and the J_β gene segment. Two 40-base oligonucleotides complementary to a portion of these putative D_β gene segments and their 5′ flanking sequence were synthesized using the technique of Horvath *et al.*[24]. The oligonucleotide T5 is complementary to a region starting at nucleotide position 56 and ending at position 95 of the published sequence of the cDNA clone 86T5 (ref. 19). T6 is complementary to the cDNA clone TM86 at positions 15–54 (ref. 19). We have hybridized these oligonucleotides to Southern blots of a genomic cosmid clone known to contain the $J_{\beta1}$ and $J_{\beta2}$ gene clusters. T6 hybridizes most strongly to a restriction fragment located to the 5′ side of the $J_{\beta2}$ gene cluster, while T5 hybridizes more strongly to the region 5′ to the $J_{\beta1}$ gene cluster (G.S., unpublished observations). We have determined the nucleotide sequence in the vicinity of the hybridizing DNA fragments and have linked them by restriction mapping and DNA sequencing to their respective J_β gene clusters (Fig. 1 and unpublished results).

Each of these hybridizing regions appears to contain a D_β gene segment by the following criteria (Fig. 1). (1) There is a 5′ recognition sequence for DNA joining with a 12-base pair spacer sequence. (2) There is a 3′ recognition sequence for DNA joining with a 23-base pair spacer. Both of these sequences are similar to their immunoglobulin counterparts (Fig. 2). (3) One of these sequences, or a highly similar D_β gene segment, appears

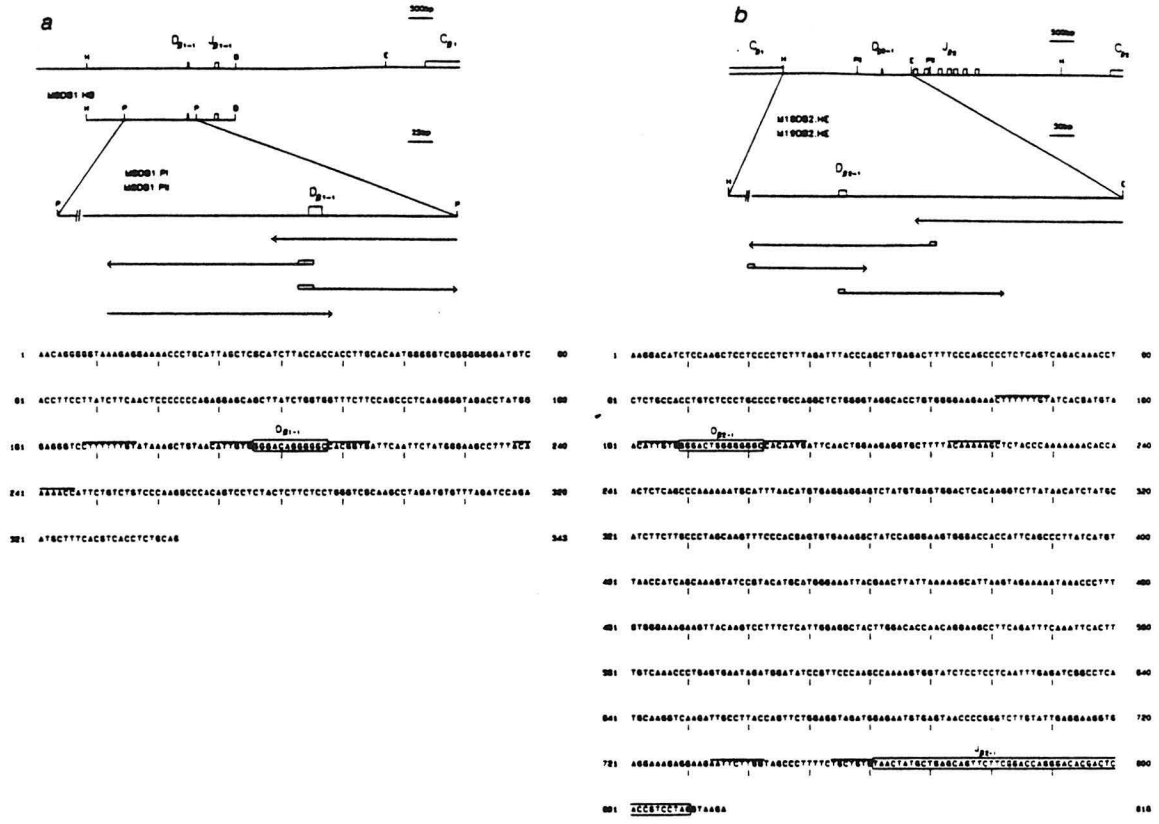


Fig. 1 Nucleotide sequences of germ-line $D_{\beta 1\text{-}1}$ and $D_{\beta 2\text{-}1}$ gene segments. *a*, The $D_{\beta 1\text{-}1}$ subclones sequencing strategy and sequence. *b*, The $D_{\beta 2\text{-}1}$ subclones, sequencing strategy and sequence. The sequence from position 410 to position 844 had been previously determined on both strands[23]. The coding regions of the D_β and J_β joining segments are boxed, and the 5′ and 3′ heptamer and nonamer recognition signals are overlined. Sequences were obtained by subcloning restriction fragments into the vectors M13mp8, M13mp18 and M13mp19 (ref. 47) and sequenced using the dideoxynucleotide method[48], as modified by E. Strauss (unpublished). Synthetic oligonucleotides complementary to mouse genomic sequences were utilized as sequencing primers in some cases. The positions of these primers are indicated by the open box at the ends of the arrows. Restriction enzyme sites are abbreviated as follows: B, *Bam*HI; E, *Eco*RI; H, *Hind*III; P, *Pst*I; PII, *Pvu*II.

to form part of the variable region of the 86T1β cDNA clone[19]. In addition, the coding sequences for these D_β gene segments are similar to some immunoglobulin D_H gene segments (Fig. 2*a*). We will denote the *D* gene segment to the 5′ side of the $J_{\beta 1}$ gene cluster as $D_{\beta 1\text{-}1}$ and the *D* gene segment to the 5′ side of the $J_{\beta 2}$ gene cluster as $D_{\beta 2\text{-}1}$. The $D_{\beta 1\text{-}1}$ gene segment is located 647 base pairs (bp) upstream from the 5′-most germ-line $J_{\beta 1}$ gene segment (Fig. 1*a*; M.K., R.H., E.S., unpublished results), while the $D_{\beta 2\text{-}1}$ gene segment is located 578 bp upstream from the 5′-most germ-line $J_{\beta 2}$ gene segment (Fig. 1*b*). Because the murine immunoglobulin heavy chain locus also has a D_H gene segment, D_{Q52}, found 696 bp 5′ to its *J* gene cluster[11], there may be some functional significance of a single *D* gene segment closely linked to the *J* gene cluster in the germline.

There probably are additional D_β gene segments that have not yet been located in the germ-line DNA. With the exception of three nucleotides missing from the 3′ end, the coding and 5′ flanking sequences of the $D_{\beta 2\text{-}1}$ gene segment are identical to the D_β gene segment expressed in the TM86 cDNA clone. We believe, therefore, that this cDNA clone includes the $D_{\beta 2\text{-}1}$ gene segment. The 86T5 cDNA D_β gene segment is also shorter than the $D_{\beta 1\text{-}1}$ gene segment by three base pairs, and in addition it has one difference at its 3′ end and one in the 5′ flanking region out of the 90 base pairs compared. It is therefore possible that the 86T5 D_β gene segment is derived from a different germ-line gene; comparable similarity between members of a family is seen in several immunoglobulin D_H gene segment families[25]. Alternatively, somatic mutation or polymorphism between inbred mouse strains could account for these differences. In addition, the D_β gene segment in 2B4.71 , a cDNA clone from a helper T-cell hybridoma[21] specific for cytochrome c, is different from the three other D_β gene segments. Thus the two D_β gene segments identified each lie to the 5′ side of their corresponding J_β gene segment clusters and it appears likely that additional D_β gene segments will lie to the 5′ side of the $J_{\beta 1}$ gene cluster.

Structure of germ-line D_β

Figure 2 shows the germ-line $D_{\beta 1\text{-}1}$ and $D_{\beta 2\text{-}1}$ coding and flanking sequences aligned with each other, with the D_β sequences from several β-chain cDNA clones, and with some representative sequences from the immunoglobulin heavy chain gene family. The $D_{\beta 1\text{-}1}$ gene segment has 12 nucleotides while the $D_{\beta 2\text{-}1}$ gene segment has 14 nucleotides (Fig. 2*a*) and they therefore resemble murine D_H segments, which range in size from 10–23 nucleotides[25]. The sequences of both the $D_{\beta 1\text{-}1}$ and $D_{\beta 2\text{-}1}$ gene segments are G-rich and they are similar at 10 out of 12 nucleotides compared; they are also similar to the presumed D_H gene segment found in the heavy chain of some oxazalone-binding immunoglobulins[26], and to the D_H region found in the heavy chain of the 18-81A-2 Abelson murine leukaemia virus (A-MuLV)-transformed cell line[27] (Fig. 2*a*). It has been

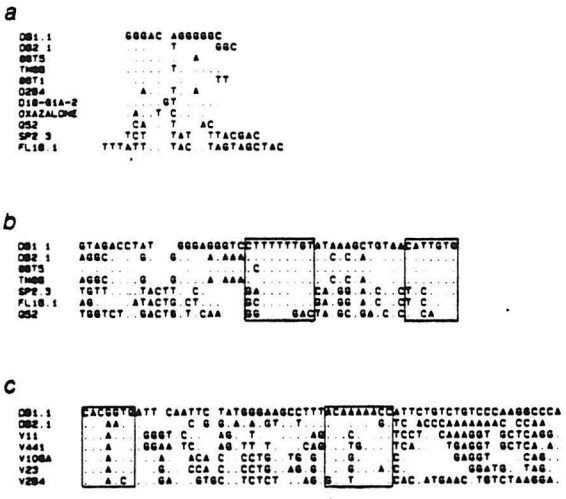


Fig. 2 A comparison of the D_β coding and flanking sequences with those found in immunoglobulin gene segments. *a*, Comparison of D_β and D_H gene segment coding regions. 86T5 and 86T1 are from β cDNA clones derived from thymus RNA[19]; TM86 is a cDNA clone from the 3.3T hybridoma[19]; D2B4 is derived from a cDNA clone from a cytochrome c-specific T-helper hybridoma[21]; oxazalone and D18-81A-2 are D_H coding sequences deduced from the sequence of immunoglobulin heavy chain mRNAs[26,40]; Q52, Sp2.3, FL16.1 are germline D_H segments[11,12,25]. *b*, Comparison of D_β and D_H gene segment 5′ flanking regions. *c*, Comparison of D_β and V_H gene segment 3′ flanking regions. The flanking sequence of four unrelated V_H gene segments that were aligned to each other are shown. The sequences are from the indicated references: V11 (ref. 49), V441 (ref. 50), V108A (ref. 51) and V23 (ref. 52). V2B4 is the V_β region gene from a β-chain cDNA clone from a cytochrome c-specific helper-T hybridoma[21]. Boxed sequences represent the nonamer and heptamer recognition signals for DNA rearrangement. Gaps were inserted to maximize homology.

previously observed that the D_β gene segment in the 2B4.71 cDNA clone is similar to the immunoglobulin D_H gene segment D_{Q52} (ref. 21). It has also been observed that the coding and flanking regions of the mouse and human $D_{\beta1\text{-}1}$ gene segments are highly conserved[28]. The significance of the similarity observed between the mouse D_β and D_H gene segments is unclear. Figure 2*b* shows the 5′ flanking regions of the $D_{\beta1\text{-}1}$ and $D_{\beta2\text{-}1}$ gene segments aligned with each other, with the 86T5 and TM86 β cDNA clones and with representative members of four murine D_H gene segment families[25]. Figure 2*c* shows the aligned 3′ flanking sequences of the $D_{\beta1\text{-}1}$ and $D_{\beta2\text{-}1}$ gene segments. These 3′ flanking sequences are also aligned with immunoglobulin V_H sequences as opposed to D_H gene segments, since the V_H gene segments also have 23-base pair spacers in their 3′ joining signals. Gaps were inserted to maximize the similarity between the sequences. The 5′ flanking sequences of the $D_{\beta1\text{-}1}$ and $D_{\beta2\text{-}1}$ gene segments are similar at 35 out of 47 positions compared. The $D_{\beta1\text{-}1}$ and $D_{\beta2\text{-}1}$ 3′ flanking sequences share 31 out of 60 nucleotides compared. Comparison of the far-flanking regions of the $D_{\beta1\text{-}1}$ and $D_{\beta2\text{-}1}$ gene segments indicates that they share the highest similarity at the recognition signals. Aside from these signals, these flanking regions have little sequence similarity with either the 5′ flanking region of D_H gene segments or the 3′ flanking region of V_H gene segments.

D_β recognition signals

As was noted by Early *et al.*[7], the 12 and 23 nucleotide spacer sequences of immunoglobulin joining recognition sequences correspond to approximately one and two turns of the DNA helix, respectively. An interesting feature of the recognition sequences for immunoglobulin V_H and V_L gene rearrangements is that the one-turn recognition sequence is always joined to a

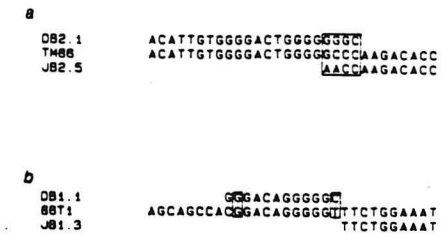


Fig. 3 Junctional and N-region diversity in β-chain rearrangement. Boxed regions indicate gene segment junctions. *a*, The cDNA clone TM86, compared with germline gene segments $D_{\beta2l}$ and $J_{\beta2\text{-}5}$. The last three nucleotides of $D_{\beta2\text{-}1}$ and the first two nucleotides of $J_{\beta2\text{-}5}$ have been deleted in the joining event, and one nucleotide of TM86 (a C at position 20) is not encoded by either germline gene segment. *b*, The cDNA clone 86T1, compared with germline gene segments $D_{\beta1\text{-}1}$ and $J_{\beta1\text{-}3}$. One nucleotide at the 5′ end of the $D_{\beta1\text{-}1}$ gene segment has been deleted or substituted in the D_β–J_β joining. One nucleotide of 86T1 (a T at position 20) is not encoded by either the $D_{\beta1\text{-}1}$ or $J_{\beta1\text{-}3}$ gene segments. No nucleotides were delected from the $J_{\beta1\text{-}3}$ gene segment during the D_β–J_β joining event; the nucleotide 5′ of the first T is a G present in the joining sequence.

two-turn recognition sequence. This rule of one- and two-turn joining is followed in the β-gene family of T-cell receptors as well (Fig. 2*b*, *c*)[21–23]. The V_β and J_β gene segments have two-turn and one-turn recognition signals, respectively, and the D_β gene segment has a one-turn recognition signal to its 5′ side and a two-turn recognition signal to its 3′ side. This organization of recognition sequences permits several types of V_β gene formation consistent with the one- to two-turn joining rule: V_β gene segments may join directly to J_β gene segments, V_β–D_β–J_β joining may occur involving a single D gene segment, or two or more D_β gene segments could join between the V_β and J_β gene segments. Thus, the organization of the recognition sequences for the β-gene family should permit additional combinatorial joining possibilities to be employed in the generation of V_β diversity.

Junctional and N-region diversity

Junctional diversity in immunoglobulin variable regions is caused both by the imprecise joining of V_L and J_L or V_H, D_H, and J_H gene segments that can lead to the deletion of nucleotides from the ends of these gene segments[10–12] and by the addition of nucleotides at the junction between V_H and D_H or D_H and J_H gene segments that are not found in either of the germ-line counterparts (N-region diversity)[11,13]. Analysis of the coding regions of germ-line D_β and J_β gene segments indicates that both mechanisms for generating junctional diversity occur in the β-chain of the T-cell receptor. Figure 3*a* compares the published sequence of the TM86 cDNA clone[19], which contains the $D_{\beta2\text{-}1}$ sequence joined to the $J_{\beta2\text{-}5}$ gene segment, to the germ-line sequences of $D_{\beta2\text{-}1}$ and $J_{\beta2\text{-}5}$. In the TM86 cDNA three bases have been deleted from the 3′ end of the $D_{\beta2\text{-}1}$ gene segment, and two bases have been deleted from the 5′ end of the $J_{\beta2\text{-}5}$ gene segment. In addition, N-region diversity is indicated by the presence of a cytidine nucleotide not found at the end of either germ-line gene segment that is present at the D–J boundary. A second possible example of both types of diversity can be found by comparing the published sequence of the 86T1 cDNA clone[19] that contains a complete variable region gene segment joined to the $D_{\beta1\text{-}1}$ and $J_{\beta1\text{-}3}$ (M.K. and R.H., unpublished) gene segments (Fig. 3*b*). Although the central ten base pairs of the 86T1 D_β gene segment are identical to the germ-line $D_{\beta1\text{-}1}$ sequence, the end nucleotides have been altered. The first nucleotide of the $D_{\beta1\text{-}1}$ gene segment at the V_β–D_β boundary has been deleted or substituted, while the last nucleotide at the D_β–J_β boundary has been replaced with a different base. In this latter example, it is also possible that no deletions or additions occur, but instead a D_β gene segment that is slightly different

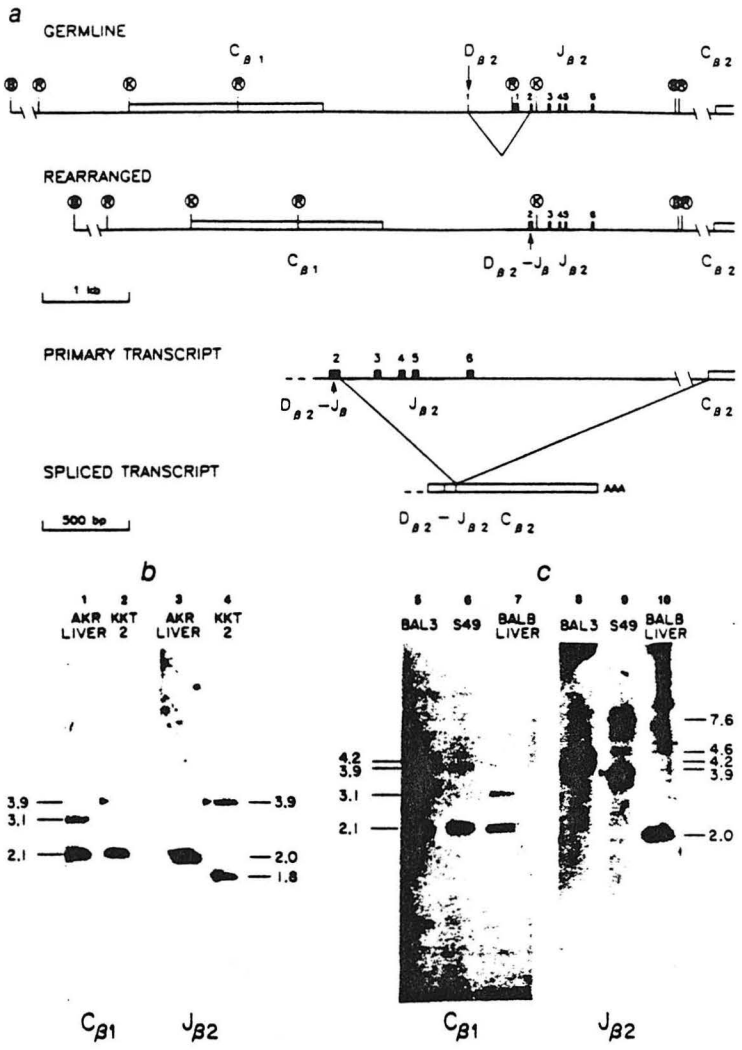


Fig. 4 D_β–J_β joining in T-cell tumour DNA. Southern blots[53] of *Eco*RI digested DNA were hybridized with a $C_{\beta 1}$ cDNA clone[54] and a fragment from a germ-line cosmid clone that contains $J_{\beta 2}$ sequences[23]. Conditions used for the Southern blots have been described previously[55]. *a*, A model for the rearrangement and expression of the partially joined (D_β–J_β) genes. The C_β genes are depicted as open boxes: the separate exons of these genes are not shown. The $D_{\beta 2}$ and $J_{\beta 2}$ gene segments are numbered and are shown as vertical lines or filled boxes. Some relevant restriction sites for the Southern blot analysis are indicated. B, *Bam*HI; K, *Kpn*I; E, *Eco*RI. The rearrangement presented in the figure joins the $D_{\beta 2\text{-}1}$ to the $J_{\beta 2\text{-}2}$ gene segment eliminating the third *Eco*RI site as well as $J_{\beta 2\text{-}1}$ and leading to the rearranged fragments described in the text. Rearrangement to other $J_{\beta 2}$ gene segments is also possible. The scale used to depict the transcripts is expanded twofold. The dotted lines at the 5′ end reflect uncertainty about the structure of this part of the transcript. *b*, Hybridization of the $C_{\beta 1}$ cDNA (lanes 1, 2) and $J_{\beta 2}$ fragment (lanes 3, 4) to AKR strain tumour KKT-2 and to AKR liver DNA. *c*, Hybridization of the $C_{\beta 1}$ (lanes 5–7) and $J_{\beta 2}$ (lanes 8-10) probes with DNA from the BALB/c tumours BAL3 and S49 and with BALB/c liver DNA. In both *b* and *c* the filter was first hybridized with $C_{\beta 1}$ the probe was removed, and the same filter was then hybridized with the $J_{\beta 2}$ probe. The arrows denote the identically-sized rearranged fragment that could be detected by both probes. The estimated sizes in kilobases of the hybridizing fragments are indicated.

from the $D_{\beta 1\text{-}1}$ gene segment may be present in 86T1. It is also possible that inter-strain sequence polymorphisms can account for these differences, since the germ-line sequences are from B10 mice and in this case the cDNA sequence is from BALB/c mice. Because the only differences are at the boundaries between the joined gene segments, we consider these latter possibilities to be less likely.

D_β–J_β joining without *V*–*D*–*J* joining

A productive DNA rearrangement is defined as one in which the *V*, *D* and *J* (or *V*–*J*) joining process occurs in such a way that a functional polypeptide chain can be expressed. In immunoglobulin genes a variety of nonproductive rearrangements may occur, including incomplete rearrangements of D_H and J_H gene segments joined together without a V_H gene segment. Such incomplete heavy chain rearrangements occur frequently in immature B cells[29,30] but are also found in some mature B cells[11,31] and T cells as well[12], and are thought to be the first step in the assembly of heavy-chain variable region genes. The β-gene segments may undergo *D*–*J* rearrangements in the absence of V_β joining in a manner similar to their immunoglobulin counterparts. We have characterized the β-gene rearrangements in a large number of T-cell lines, hybridomas, and tumours (M.K., J. Goverman, M. Malissen, R.H., manuscript in preparation) and have observed several cases in which the DNA from T-cell tumours have undergone an apparent D_β–J_β joining. For example, when germ-line DNA is digested with the restriction enzyme *Eco*RI and hybridized with a $C_{\beta 1}$ cDNA probe, there are two hybridizing fragments of 2.1 and 3.1 kilobases (kb). The 3.1 kb fragment is at the 3′ end of the C_β gene beginning close to the translational stop codon and ending just 5′ to the first $J_{\beta 2}$ gene segment[23] (Fig. 4*a*). In the T-cell tumours S49, BAL3, and KKT-2, the 3.1 kb fragment at the 3′ end of the $C_{\beta 1}$ gene is replaced by a larger fragment that is either 3.9 kb (S49 and KKT-2) or 4.2 kb (BAL3) (Fig. 4*b*, *c*). Hybridization of the same filters with the $J_{\beta 2}$ probe reveals two rearranged fragments in each case, including one fragment identical in size to the one detected with the $C_{\beta 1}$ probe. Southern blots of DNA from these tumours were generated from DNA digested with *Bam*HI and hybridized with the $C_{\beta 1}$ and $J_{\beta 2}$ probes. Digestion of germ-line DNA with this enzyme will generate a single fragment that should hybridize with both probes (Fig. 4*a*). In the DNA from each of the tumours one of the fragments hybridizing with the $J_{\beta 2}$ probe and the only fragment hybridizing with the $C_{\beta 1}$ cDNA are identically sized and approximately 1,000 bp smaller than the germ-line fragment (M.K., R.H., unpublished observations). We therefore conclude that the rearrangements 3′ to the $C_{\beta 1}$ gene in these tumours are small deletions. Since this deletion spans the *Eco*RI site just to

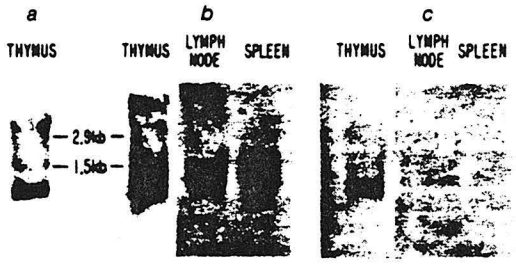


Fig. 5 Hybridization of D_β and C_β probes to RNA from lymphoid tissues. The probes are as follows: *a*, T6 synthetic oligonucleotide; *b*, C_β cDNA clone; *c*, T5 synthetic oligonucleotide. RNA was prepared from thymus, spleen, and lymph node of 4-week-old DBA 1/J female mice by homogenization of the tissues in guanidinium thiocyanate followed by centrifugation through a cushion of CsCl[56]. Northern blots of total cell RNA and hybridization with a C_β probe were performed as described previously[57]. Hybridization with oligonucleotides was done at room temperature with end-labelled synthetic probe at a concentration of 0.5 pmol ml^{-1}. The blots were washed in 2×SSC/0.1% SDS at room temperature followed by washing in the same solution at 37 °C. One filter was used in *a*, *b* and *c*; following hybridization and exposure to film, the probe was removed by stringent washing before retesting with a second probe. The migration distance of *Escherichia coli* ribosomal RNA molecular weight markers is indicated. Other markers included mouse 28+18S ribosomal RNA (4.8 and 1.9 kb) and MS2 phage RNA (3.6 kb).

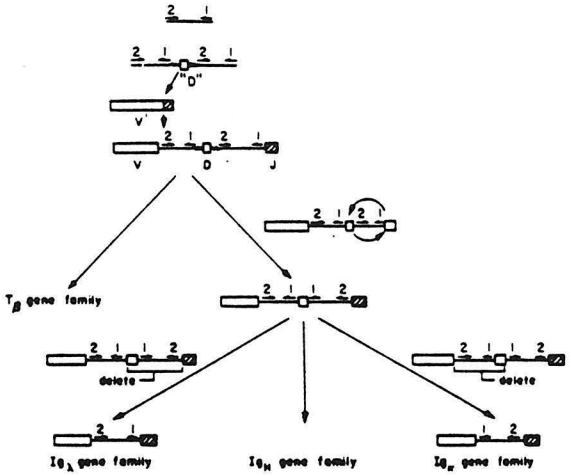


Fig. 6 A model for the evolution of the immunoglobulin and β T-cell receptor gene families. The number 1 represents the one-turn recognition signal; the number 2 represents the two-turn recognition signal.

the 5′ side of the $J_{\beta 2}$ gene cluster, the most likely explanation for these events is that the $D_{\beta 2\text{-}1}$ gene segment has joined to a $J_{\beta 2}$ gene segment (Fig. 4*a*). The slight difference in the size of the rearranged fragment in different tumours could reflect joining to different $J_{\beta 2}$ gene segments. We have similar evidence that $D_{\beta 1\text{-}1}$ to $J_{\beta 1}$ joining events can occur as well (unpublished data). Thus it appears likely that the $D_{\beta 1\text{-}1}$ and $D_{\beta 2\text{-}1}$ gene segments do participate in both complete (V_β–D_β–J_β) and partial (D_β–J_β) DNA rearrangements. By analogy with immunoglobulin heavy-chain genes, these partial rearrangements may represent the first step in β-chain variable region gene assembly.

D_β–J_β transcription not using V_β promoters

Northern blot analyses have been carried out on RNA from thymus, lymph node, and spleen using both a C_β probe (Fig. 5*b*) and the oligonucleotides T5 and T6 (Fig. 5*a*, *c*). The C_β probe detects two species, approximately 1.3 and 1.0 kb in size, when hybridized to total RNA from the thymus. Similar results have been obtained with poly(A)$^+$ RNA (M.K., G.S., unpublished observations). Spleen and lymph node RNA have a single C_β hybridizing transcript of 1.3 kb. When RNA from these three lymphoid tissues is hybridized with either the T5 or T6 oligonucleotide, only the 1.0 kb thymus RNA is detected. It is important to note that both T5 and T6 are complementary to 6 bases of D_β coding sequences and 34 bases of 5′ flanking sequence. These two probes therefore do not hybridize with fully rearranged V_β–D_β–J_β genes since these rearrangements eliminate the 5′ D_β flanking sequence. Because the probes also do not cross-hybridize with each other, we conclude that at least two D_β gene segments contain 5′ promoters that permit expression in thymus: one is the $D_{\beta 2\text{-}1}$ gene segment and the second is the $D_{\beta 1\text{-}1}$ gene segment or a highly similar *D* gene segment such as the one found in the 86T5 cDNA clone. It has recently been demonstrated that each immunoglobulin D_H gene segment also has a promoter located in the 5′ flanking DNA[32].

We have hybridized both the C_β probe and the D_β oligonucleotide probes to RNA from the S49 and KKT-2 tumours that are believed to have joined $D_{\beta 2}$–$J_{\beta 2}$ gene segments. The KKT-2 tumour has only a 1.0 kb transcript that hybridizes with both the T6 D_β oligonucleotide and a probe specific for the 3′ end of $C_{\beta 2}$. Several other T-cell tumours and T-cell hybridomas contain 1.0 kb transcripts that hybridize with both C_β probes and the T6 oligonucleotide. However, RNA from S49 does not hybridize with either oligonucleotide, and it contains a 1.3-kb but not a 1.0-kb transcript when hybridized with the C_β cDNA probe (M.K., unpublished observations).

Based on the blot-hybridization and nucleotide sequence data presented here and the partial nucleotide sequences of the TM86 and 86T5 cDNA clones, we have presented a model for the rearrangement and expression of the D_β gene segments and their 5′ flanking sequences (Fig. 5*a*). D_β–J_β rearrangement may stimulate transcription from a promoter located 5′ to several of the D_β gene segments. This transcript is 1.0 kb long, polyadenylated, and contains a joined D_β–J_β gene segment spliced to the exons normally used in the C_β gene. Although we cannot formally rule out transcription from this promoter in the absence of gene rearrangement, it should be noted that the KKT-2 tumour has no germ-line $J_{\beta 2}$ clusters (Fig. 4*c*) and that a transcript derived from a promoter 5′ to the $D_{\beta 2\text{-}1}$ gene segment that includes the intron between $D_{\beta 2\text{-}1}$ and $J_{\beta 2\text{-}1}$ as well as the exons of $C_{\beta 2}$ would be at least 1.6 kb. No transcript of this length has been detected with any of the probes employed. The prevalence of the 1.0 kb transcript in thymus, as opposed to lymph node and spleen, would suggest that many of the immature cells in this tissue have undergone a partial (D_β–J_β) rearrangement. However, as some functional mature T lymphocytes have a 1.0-kb transcript that hybridizes with the T6 oligonucleotide (M.K., unpublished), there is probably a low level of 1.0-kb transcript in lymph node and spleen that could not be detected on our blots.

Immunoglobulin gene expression in B lymphocytes is thought to result from gene rearrangements that bring into proximity two widely separated elements: a promoter sequence located 5′ to the *V* gene segments[33,34] and an enhancer sequence located in the intron between the *J* gene segments and the *C* genes[35–39]. In the case of the D_β–J_β transcript, however, the location of the promoter relative to any postulated enhancer would not be greatly affected by the relatively small deletions involved in the D_β–J_β joining we have described. Therefore the means by which D_β–J_β joining might stimulate transcription are unknown. Furthermore, Northern blots of RNA from the S49 tumour indicate that rearrangement is necessary but may not be sufficient for transcription. Currently, we are further characterizing the D_β–J_β

DNA rearrangements and D_β transcripts in a number of T-cell tumours in order to understand better the relationship between the rearrangements and the 1.0 kb C_β transcripts that we have described.

Are D_β–J_β transcripts translated?

In the immunoglobulin gene families, transcripts have been described that originate from unrearranged κ[40] and heavy-chain genes[41] or from incompletely rearranged (D_H–J_H) heavy-chain gene segments[27,41]. Some of the RNA molecules do not give rise to detectable polypeptides and are referred to as sterile transcripts[27,42]. These sterile transcripts might not be physiologically significant, or else the RNA molecule itself might have some regulatory function. Similar speculations apply to the 1.0-kb transcripts containing D_β gene segments. In addition, it is also possible that these D_β–J_β transcripts encode functional polypeptides. If the 1.0-kb transcripts use the entire C_β gene and the same 3′ end as is found in the 1.3-kb C_β RNA, then about 100–200 nucleotides remain for the D_β coding region and its 5′ flanking sequence. The nucleotide sequences flanking the $D_{\beta 1\text{-}1}$ and $D_{\beta 2\text{-}1}$ gene segments have open reading frames for more than 50 codons. A methionine codon that is in frame with the germ-line D_β gene segments is located 5′ to both $D_{\beta 1\text{-}1}$ and $D_{\beta 2\text{-}1}$ (data not drawn for $D_{\beta 2-1}$). Moreover, the $D_{\beta 1\text{-}1}$ flanking sequence has a signal sequence for membrane insertion of a nascent polypeptide[43] 136 nucleotides 5′ to the D_β gene segment (Fig. 1*b*). A possible signal sequence is also located following the methionine codon 5′ of $D_{\beta 2\text{-}1}$ (data not shown). Alt and co-workers have demonstrated that most immunoglobulin D_H gene segments have open reading frames in their 5′ flanking sequence that begin with a methionine codon followed by a hydrophobic leader sequence. They have also demonstrated that transcripts derived from partial D_H–J_H rearrangements give rise to truncated C_μ-containing polypeptides[32]. We therefore believe that the D_β–J_β transcripts will in at least some cases give rise to truncated β polypeptides. The function, if any, of these truncated polypeptides is unknown. Because it has been proposed that the synthesis of complete heavy chain and light chain polypeptides derived from productively rearranged genes prevents further *V* gene rearrangement[40,44], we might speculate that the truncated C_μ and C_β polypeptides will in a similar fashion somehow stimulate or facilitate the *V* gene rearrangement process. It is also possible that these truncated proteins provide a stimulatory signal for cell growth similar to that provided by antigen bound to the intact receptor. Alternatively they might function at the cell surface with regard to the development of the repertoire of receptors via network interactions.

Evolution of gene rearrangements

It has been proposed that the ability of the antibody and T-cell receptor gene families to undergo DNA rearrangements evolved from the insertion of a transposable element into a primordial uninterrupted *V* gene exon[6,45] (Fig. 6). This theory proposed the existence of a transposon with asymmetric inverted repeats similar to those found in the joining sequences of the contemporary antigen-receptor gene families. If inserted in the genome, duplication of this transposon along with a short stretch of flanking sequence would lead to a more complex transposon containing a proto-*D* gene segment (Fig. 6). Subsequent insertion of this transposon into a primordial *V* gene would generate a gene segment organization, at least with respect to joining sequences, that is similar to that of the β genes. Duplication of this locus would lead to two separate gene families: the β-chain gene family and the immunoglobulin gene family. As shown in Fig. 6, an inversion bounded by the two right-hand inverted repeats could generate the precursor of the heavy chain gene family. As outlined previously[45], further duplications of this heavy chain gene precursor could have occurred to create precursors of all three immunoglobulin gene families. A deletion on the right hand or 3′ end of the transposon as indicated in Fig. 6 generates the proto-λ gene family with its organization of joining signals. This may have occurred via homologous recombination between these signals. A similar deletion at the 5′ end could give rise to the κ gene family.

This model can account for some of the unique features of variable region genes. The known genes encoding antigen receptors on B and T lymphocytes, cell-surface molecules of the major histocompatibility complex (MHC), β_2-microglobulin and some other cell-surface proteins associated with the immune system, are believed to have evolved in part from a primordial exon that encoded an immunoglobulin-like domain[46]. With the exception of the variable region genes, each domain in these proteins having an immunoglobulin structure is encoded by a single exon. The insertion of a transposon into a primordial variable region gene can account for the fact that all the *V* gene families are split into several exons at a similar location. The theory accounts for the presence of a *D* gene segment in only two of the four antigen-receptor gene families and also explains the different configurations of DNA rearrangement signals in these families. While the evolutionary history outlined above implies that the β-chain gene family arose first, several other pathways beginning with insertion of a slightly different transposon are also possible[45]. Finally, this theory proposes a relatively simple explanation for the evolution of the DNA rearrangement signals in that it proposed these signals evolved from the inverted terminal repeats of a mobile DNA sequence but it remains unclear how these signals and the proteins that recognize them became adapted for the highly regulated rearrangements of the immune system.

This work was supported by grant AI 17565 from the NIH and by a grant from T-cell Sciences, Inc. We thank Dr Irving Weissman, Stanford University, for providing T-cell tumours, Dr Roger Perlmutter for helpful and stimulating discussions, and Dr Suzanna Horvath and Marilyn Seibel for synthesis of the oligonucleotides. We also thank Tim Hunkapiller, Astar Winoto, Drs Rick Barth, Joan Goverman, Joan Kobori, Bernard Malissen, Marie Malissen and Martha Zuniga for critique of the manuscript, Dr Marie Malissen for providing the cosmid clone and plasmid subclones, Debbie Maloney for technical assistance, and Bernita Larsh for preparation of the manuscript.

Note added in proof: While this paper was in the press, the sequence of the $D_{\beta 1-1}$ gene segment and the presence of this gene segment in the 86T1 and 86T5 cDNA clones was reported by Kavaler *et al.*[58].

Received 29 June; accepted 23 July 1984.

1. Early, P. & Hood, L. in *Genetic Engineering* (eds Setlow, J. K. & Hollaender, A.) 157–188 (Plenum, New York, 1981).
2. Tonegawa, S. *Nature* **302**, 575–581 (1983).
3. Honjo, T. *A. Rev. Immun.* **1**, 499–528 (1983).
4. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15**, 1–14 (1978).
5. Seidman, J., Max, E. & Leder, P. *Nature* **280**, 370–375 (1979).
6. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
7. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. *Cell* **19**, 981–992 (1980).
8. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286**, 676–683 (1980).
9. Schilling, J., Clevinger, B., Davie, J. M. & Hood, L. *Nature* **283**, 35–40 (1980).
10. Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J. & Hood, L. *Nature* **283**, 497–499 (1980).
11. Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. *Nature* **290**, 562–565 (1981).
12. Kurosawa, Y. *et al. Nature* **290**, 565–570 (1981).
13. Alt, F. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4118–4122 (1982).
14. Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3450–3454 (1979).
15. Allison, J. P., McIntyre, B. W. & Bloch, D. *J. Immun.* **129**, 2293–2300 (1982).
16. Haskins, K. *et al. J. exp. Med.* **157**, 1149–1169 (1983).
17. Meuer, S. C. *et al. J. exp. Med.* **157**, 705–719 (1983).
18. Hedrick, S. M., Cohen, D. I., Nielsen, E. A. & Davis, M. M. *Nature* **308**, 149–153 (1984).
19. Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I. & Davis, M. M. *Nature* **308**, 153–158 (1984).
20. Yanagi, Y. *et al. Nature* **308**, 145–153 (1984).
21. Chien, Y., Gascoigne, N., Kavaler, J., Lee, N. & Davis, M. *Nature* **309**, 322–326 (1984).
22. Siu, G. *et al. Cell* **37**, 393–401 (1984).
23. Malissen, M. *et al. Cell* **37**, 1101–1110 (1984).
24. Horvath, *J. biol. Chem.* (submitted).
25. Kurosawa, Y. & Tonegawa, S. *J. exp. Med.* **155**, 201 (1982).
26. Kaartinen, M., Griffiths, G., Markham, N. & Milstein, C. *Nature* **304**, 320–324 (1983).
27. Alt, F. W., Rosenberg, N., Enea, V., Siden, E. & Baltimore, D. *Molec. cell. Biol.* **2**, 386–400 (1982).
28. Clark, S. *et al. Nature* **311**, 387–389 (1984).
29. Alt, F. Rosenberg, N., Lewis, S., Thomas, E. & Baltimore, D. *Cell* **27**, 381–390 (1981).
30. Nelson, K., Haimovich, J. & Perry, R. *Molec. cell. biol.* **3**, 1317–1332 (1983).
31. Early, P., Nottenburg, C., Weissman, I. & Hood, L. *Molec. cell. Biol.* **2**, 829–836 (1982).
32. Reth, M. & Alt, F. *Nature* (in the press).

33. Kataoka, T., Nishido, T., Miyata, T., Morlwaki, K. & Honjo, T. *J. biol. Chem.* **257**, 277–285 (1982).
34. Clarke, C. *et al. Nucleic Acids Res.* **10**, 7731–7749 (1982).
35. Rice, D. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7862–7865 (1982).
36. Oi, V., Morrison, S., Herzenberg, L. & Berg, P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 825–829 (1983).
37. Gillies, S., Morrison, S., Oi, V. & Tonegawa, S. *Cell* **33**, 717–728 (1983).
38. Queen, C. & Baltimore, D. *Cell* **33**, 741–748 (1983).
39. Banerji, J., Olson, L. & Schaffner, W. *Cell* **33**, 729–740 (1983).
40. Perry, R. P. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 1937–1941 (1980).
41. Kemp, D. J., Harris, A. W. & Adams, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7400–7404 (1980).
42. Walker, I. D. & Harris, A. W. *Nature* **288**, 290–293 (1980).
43. Silhavy, T. S., Benson, S. A. & Emr, S. D. *Microbiol. Rev.* **47**, 313–344 (1983).
44. Alt, F. W., Enea, V., Bothwell, A. L. M. & Baltimore, D. *Cell* **21**, 1–12 (1980).
45. Hood, L. E., Hunkapiller, T. & Kraig, E. in *Modern Cell Biology* (ed. McIntosh, J. R.) 305–328 (Liss, New York, 1983).
46. Williams, A. F. *Nature* **308**, 108–109 (1984).
47. Messing, J. & Vieira, J. *Gene* **19**, 269–276 (1982).
48. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161–178 (1980).
49. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. *Cell* **25**, 59–66 (1981).
50. Ollo, R., Auffray, C., Sikorav, J.-L. & Rougeon, F. *Nucleic Acids Res.* **9**, 4099–4108 (1981).
51. Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. & Cohen, J. B. *Nature* **292**, 426–430 (1981).
52. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. *Cell* **24**, 625–637 (1981).
53. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
54. Caccia, N. *et al. Cell* **37**, 1091–1099 (1984).
55. Steinmetz, M. *et al. Nature* **300**, 35–42 (1982).
56. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
57. Kronenberg, M., Davis, M. M., Early, P. W., Hood, L. E. & Watson, J. D. *J. exp. Med.* **152**, 1745–1761 (1980).
58. Kavaler, J., Davis, M. M. & Chien, Y. *Nature* **310**, 421–424 (1984).

CHAPTER SIX

THE MOLECULAR GENETICS OF THE T-CELL ANTIGEN RECEPTOR AND T-CELL ANTIGEN RECOGNITION

Annual Review of Immunology, in press

THE MOLECULAR GENETICS OF THE T-CELL ANTIGEN RECEPTOR AND T-CELL ANTIGEN RECOGNITION

Mitchell Kronenberg, Gerald Siu, Leroy E. Hood and Nilabh Shastri

Division of Biology, California Institute of Technology

Pasadena, California 91125

INTRODUCTION

T and B lymphocytes participate in the vertebrate immune response through the recognition and elimination of foreign pathogens and macromolecules. Lymphocytes possess both the specificity and the ability to react with a broad range of structures, properties that are mediated through both T-cell antigen receptors and immunoglobulins that serve as antigen receptors for B cells. Despite some similarities, T-cell antigen recognition and function is different in several important respects from the corresponding processes in B cells. First, T cells only recognize antigens that are present on the surfaces of other cells in the context of polymorphic cell surface molecules encoded by the major histocompatibility complex (MHC) (1-3). The mechanism by which T-cell receptors express an apparent dual specificity for both antigen and polymorphic determinants of the MHC-encoded molecule is not known and has been the subject of much debate. Second, T cells can be subdivided into separate functional categories including cytotoxic effector cells (T_C) (4), inducer or helper cells (T_H) (5, 6) and suppressor cells (T_S) (7). Since immunoglobulins encode both antigen recognition and effector functions within the same molecule, an important question is whether antigen receptors expressed by different T-cell functional classes are distinct. Third, T cells mature in the thymus, an organ composed almost entirely of developing T lymphocytes, while B cells in mammals differentiate in the fetal liver or bone marrow (8-10). Development in the thymus is characterized by both rapid proliferation and cell death (reviewed in 11, 12), processes that are probably related to the expression and selection of T-cell receptors. Finally, a number of accessory molecules play an important role in T-cell antigen recognition and/or activation; similar molecules have not been identified for B cells. For example, in humans the T3 molecule, composed of the δ, γ and ε polypeptides, is intimately associated with the T-cell receptor and may

serve as an ion channel that functions in T-cell recognition and activation (13, 14). In addition, the human T4 and T8 molecules, and their mouse equivalents L3T4 and Lyt-2, may bind nonpolymorphic determinants on class II and class I MHC molecules, respectively, thereby stabilizing the interaction of T cells with their targets (15-19). Although much has been learned about the functional characteristics of T cells, until recently little was known about the molecular nature of the antigen-specific receptor.

The first breakthrough in the characterization of the T-cell antigen receptor came with the generation of monoclonal antibodies that could bind to only one of a panel of T-cell clones (20-26). These clone-specific antibodies were capable of affecting antigen-specific activation, suggesting that they recognized the antigen receptor. Using these serological reagents, it was demonstrated that the receptor is composed of a disulfide-linked heterodimer composed of two subunits, denoted α and β, each with a molecular weight of approximately 40-50 kdaltons. Further characterization of this heterodimer demonstrated that each chain contains portions that are variable and portions that are constant between different T cells, suggesting the presence of both α and β chain variable and constant regions (23, 27, 28).

Study of the molecular genetics of the T-cell antigen receptor began when cDNA clones encoding the α and β chains, as well as γ, a third T-cell specific gene family whose function is unknown, were isolated using subtractive or differential hybridization techniques (29-33). Just 16 months ago, the first papers that characterized cDNA clones encoding the β polypeptide of the T-cell antigen receptor were published (29, 30, 34). Since that date, over 70 papers concerned with the T-cell receptor α and β, and the γ genes of mouse and man have been written. Progress has been rapid, and in some areas the depth of knowledge of these genes already approaches that of the immunoglobulin genes. This article

summarizes the current understanding of the organization, rearrangement, ontogeny of expression and diversification of the genes encoding the α and β chains of T-cell antigen receptor. We discuss the γ genes as well, because of their obvious structural similarity to the α and β T-cell receptor genes. We also attempt to integrate the available knowledge of T-cell receptor genes with theories that account for the mechanisms responsible for MHC-restricted antigen recognition and immune response gene defects, areas that are still characterized by controversy and speculation.

STRUCTURE AND ORGANIZATION

Structure of the α, β and γ Chains

The polypeptides encoded by the α, β, and γ genes have significant structural homology to the immunoglobulin genes (29, 31-37). Many of the detailed structural features of these polypeptides have been inferred from the translated sequences of cDNA clones. The α, β and γ chains are 30-37 kdaltons prior to both glycosylation and cleavage of the leader sequence. They can each be subdivided into seven regions: a hydrophobic leader region of 18-29 amino acids that is characteristic of all cell-surface and secreted proteins, a variable (V) segment of 88-98 amino acids, a joining (J) segment of approximately 14-21 amino acids, a sequence of 87-113 amino acids that resembles an immunoglobulin constant (C) region, a connecting peptide of varying length, a transmembrane region of approximately 20-24 amino acids, and a small cytoplasmic region of 5-12 amino acids (Figure 1). The V and J segments together constitute the V regions which display significant structural similarity with immunoglobulin V regions, including a length of approximately 110 amino acids, a centrally positioned disulfide bridge spanning 63-69 amino acids, and several other conserved amino acids that are believed to be important for protein structure. The α and β chains both have a

cysteine located near their carboxyl terminus in the connecting peptide that probably participates in the known disulfide bond linking the α and β subunits of the heterodimer. The γ chain also has a similar cysteine in the connecting peptide that may participate in a disulfide bond with another γ chain to form a homodimer or possibly with another polypeptide to form a heterodimer. The β and γ chain have one, and the α chain has two positively charged amino acids in its transmembrane region. The presence of charged amino acids in this hydrophobic stretch is unusual, although not unprecedented (38). One or more of the positive charges in α and β could form an ionic interaction with a negatively-charged aspartic acid in the transmembrane portion of the δ chain of the T3 complex or possibly with other membrane proteins.

Chromosomal Locations of the α, β and γ Genes

The chromosomal locations of the murine and human α, β, and γ loci have been determined by a variety of techniques. The results of these studies are summarized in Table 1. In mice, each of the three T-cell gene families that undergo rearrangement is encoded on a separate chromosome. The gene that encodes the murine C_α region is located on chromosome 14, bands C or D (39, 40). A group of related V_α region genes were shown to be linked to the C_α gene (40). In the human α gene family (41) and in the mouse (42, 43) and human (44-46) β gene families as well, all the data are consistent with chromosomal linkage of V and C genes. Thus, as in the three unlinked immunoglobulin gene families, V and C genes that are members of the same gene family yet separate in the germline (see below), are chromosomally linked.

The genes that encode the C_β regions in mice are located on chromosome 6, band B (45, 47), which also contains the immunoglobulin κ gene family (48, 49) and the gene encoding the Lyt-2 molecule expressed primarily by MHC class I-specific of T cells (50, 51). Genetic analysis using monoclonal antibodies directed against

V_β genes in one case (42, 52, 53), or a combination of both monoclonal antibodies and restriction fragment length polymorphisms in another (43), have shown that the β genes are approximately 10-13 centimorgans from the very closely linked Lyt-2 and C_κ genes. The murine γ genes have been located to band A2 or A3 of chromosome 13 (40).

The chromosomal locations of the human α, β and γ loci have also been determined. As described for the mouse, in humans the three T-cell rearranging gene families are unlinked (Table I). The α locus is located on chromosome 14, bands q11-12 (41, 54-57). The immunoglobulin heavy chain genes are also on chromosome 14, but they are located at band q32 (58) and therefore are not closely-linked to the α genes. The region of chromosome 14 containing the α genes is often rearranged in T-cell malignancies (59-61). Although there is so far no proof that these rearrangements directly involve either the α gene locus or oncogenes, they may be analogous to the proto-oncogene rearrangements observed in the immunoglobulin loci in B lymphocyte malignancies (62, 63).

The relatively close linkage of the β and κ chain genes in mice is not an essential feature of the organization of these gene families. The human β chain genes have been localized to chromosome 7 (44-46, 64-66) and the human κ genes are found on chromosome 2 (67). Data from *in situ* hybridization studies indicate that the β genes are located on bands q32—35 of chromosome 7 (46, 65, 66), although in some individuals a significant amount of hybridization is observed on the short arm of chromosome 7 (bands p15-21) (46). In the earliest study in which metaphase chromosome spreads from a single individual were analyzed, most of the observed grains were in fact located around 7p13-21, and relatively few were located near 7q32-35 (45). There is no simple explanation for the variable hybridization of β probes to the short arm of chromosome 7. Although the human γ genes are located at chromosome 7 band p15 (41), the β and γ genes have

diverged extensively and should not cross hybridize. Chromosomal rearrangements with breakpoints that are close to the α, β and γ gene loci are common in lymphocytes from ataxia telangiectasia patients (68). To a much lesser extent, such abnormalities are also seen in mitogen-stimulated T lymphocytes from normal individuals (69, 70), suggesting that fragile sites in these chromosomes are related to the sites that undergo normal gene rearrangements in T lymphocytes.

Organization of the α, β and γ Gene Families

Using cDNA clones as probes, the organization of the α, β and γ gene families has been characterized (Figure 2). As in the immunoglobulins, the genes that encode these molecules are actually composed of two parts: a variable gene, and a constant gene. The variable gene is composed of either two (V and J) or three (V, D and J) gene segments (71-76). Each gene family has multiple V and J gene segments and one to three constant genes (71-79). The V, D and J gene segments are separate in the germline and are brought together by DNA rearrangement during T-lymphocyte differentiation to form the complete V gene (71-76). This DNA rearrangement is mediated by rearrangement signals located directly 3' to the V gene segments, 5' to the J gene segments, and on either side of the D gene segments (71-78, 80-82) (Figure 3). The sequences of these rearrangement signals in the three families are similar to one another and to those found in the immunoglobulin gene families. These sequences are composed of a conserved heptamer, 5' CACAGTG 3', and a conserved A/T-rich nonamer, separated by a nonconserved spacer sequence of either 12 or 23 base pairs (Figure 3). The rearrangement signals of any two gene segments that can undergo joining are, with the exception of the spacer sequence, nearly inverse complements of one another and therefore these sequences on the same DNA strand can hypothetically base pair with one another (Figure 3a). As in immunoglobulin genes (83-85), DNA rearrangement is believed to occur only when

one gene segment has a rearrangement signal with a 12 base pair spacer (approximately one turn of the DNA helix and therefore denoted a one-turn signal) and the other gene segment has a 23 base pair spacer (approximately two turns of the DNA helix and designated a two-turn signal) (Figure 3b).

Variable Gene Segments

The germline variable (V) gene segments of the α, β and γ gene families are composed of two exons separated by an intron of 100-400 base pairs. The first exon is approximately 49 base pairs in length and encodes most of the hydrophobic leader sequence, while the second exon of approximately 300 base pairs encodes the remaining five amino acids of the leader and the V segment which comprises the first 88-98 amino acids of the variable region gene (71-76). All of the α, β and γ V gene segments have a two-turn signal for DNA rearrangement immediately adjacent to their 3' ends, as do the immunoglobulin V_H and V_λ gene segments (85) (Figure 3). The rearrangement of a particular V_β gene segment deletes others from the same chromosome, suggesting that many V_β gene segments are located 5' of both the deleted V_β gene segments and the C_β genes (R. Barth and J. Goverman, unpublished data). However, the murine β locus has V_β gene segments distributed both 5' and 3' of the C_β genes. One V_β gene segment is located 10 kb 3' to $C_{\beta 2}$ in reverse transcriptional orientation (Figure 2) and is utilized by a functional T-cell clone (86). It is not known if other V_β gene segments are located 3' of the C_β genes. The orientation of the V_α and V_γ gene segments, as well as the V_H and V_L gene segments, to their respective C genes is also unknown.

Figure 4 shows the amino acid sequence of the 16 known murine V_β gene segments. Like the immunoglobulin V gene families, the T-cell antigen receptor V gene families can be divided into different subfamilies, each consisting of closely related V gene segments more than 75% similar in DNA sequence (87-89). The 16 mouse V_β sequences constitute 14 different subfamilies. We have proposed an

arbitrary but simple numerical nomenclature for the V_β gene segment subfamilies and gene segments (88) which is presented in Figure 4. According to this nomenclature, members of the same subfamily share the first digit and differ in the second; therefore $V_{\beta 8.1}$, $V_{\beta 8.2}$ and $V_{\beta 8.3}$ are all members of the $V_{\beta 8}$ subfamily. The mouse V_β gene family is unique in that hybridization of Southern blots with V_β gene segment probes has shown that it consists of a large number of single-member gene subfamilies. Of the 14 V_β subfamilies, 12 are single copy, and the remaining two have only three members each (87-89). One of the three member subfamilies is $V_{\beta 8}$, and the other is $V_{\beta 5}$, although only one of the three members of the $V_{\beta 5}$ subfamily has so far been cloned and sequenced. In contrast, the murine V_κ and V_H gene segment subfamilies contain four to 50 or more members (90-92). Although the sizes of the V_β subfamilies are small, there are a larger number of known subfamilies than found in murine immunoglobulin genes, where only seven V_H subfamilies and eight V_κ subfamilies have been identified (90-92). It is not yet known whether human V_β gene segment subfamilies are organized similar to those of the mouse. At present only one human subfamily with five members has been characterized (93).

A compilation of the known murine V_α gene segment sequences is presented in Figure 5 (31, 33, 94, 95). The amino acid sequences shown were translated from the nucleotide sequence of 22 different cDNA clones. The 22 sequences can be grouped into 11 subfamilies, and Southern blots have shown these families consist of one to 10 members each (94, 95). The V_α gene family is therefore similar to the V_κ and V_H gene families in that, with one exception, all the V_α gene segment subfamilies have multiple members. A numerical nomenclature for the mouse V_α gene segment subfamilies has been proposed (94), as was done for the V_β genes (Figure 5). The murine V_γ family contain at least one subfamily of three members, only one of which appears to be utilized frequently (73) (see below).

Two of the three members of this V_γ gene segment subfamily have been linked and are in opposite transcriptional orientations (73).

Comparisons of the different V_α or V_β gene segments reveal that they can be quite diverse, differing from other gene segments in the same gene family by as much as 70% in DNA sequence and 84% in protein sequence (87-89, 93-95). This is somewhat more diverse than what is observed in the immunoglobulin V gene families, where the most divergent V gene segments of the same family differ by as much as 76% in protein sequence (90-92). Despite this diversity, the V_β and V_α genes both have conserved amino acids that are also conserved in immunoglobulin V gene segments (see below).

Diversity Gene Segments

Two murine D_β gene segments have been characterized (Figure 2). The more 5' gene segment, $D_{\beta 1}$, is 12 base pairs long and is identical in sequence in mice and humans (80-82), and the more 3' gene segment, $D_{\beta 2}$ is 14 nucleotides long (82). The sequences of both murine D_β gene segments are G-rich and similar to each other. The D_β gene segments have one-turn recognition signals for DNA rearrangement in their 5' flanking regions and two-turn recognition signals in their 3' flanking region (Figure 3) and each is located approximately 500-600 nucleotides upstream to a cluster of J_β gene segments (80-82).

No germline D_α and D_γ gene segments have been isolated and it is not known whether such gene segments exist. Sequence analysis of both germline and rearranged α and γ gene segments shows that generally all but a few nucleotides in the V-J junctional regions are encoded by the germline gene segments (73-76). These additional nucleotides could be encoded by D gene segments, or they could be the result of other mechanisms that add random nucleotides to the junctional region in the process of joining the gene segments (see below).

Joining Gene Segments

In the mouse there are 14 J_β gene segments, at least three J_γ gene segments and perhaps more than 50 J_α gene segments (71, 73-78) (Figure 2). All of the functional T-cell receptor J gene segments have one-turn recognition signals for DNA rearrangement in their 5' flanking regions (Figure 3) and a splice-donor signal at their 3' boundary that in the primary transcript permits splicing of the joined V-J to the C region sequences. The J_β, J_α and J_γ gene segments contain 15-17, 19-21 and 19 codons respectively, as well as parts of two other codons at their 5' and 3' ends (Figure 6). As in the variable gene segment, the J_α, J_β and J_γ gene segments encode several conserved amino acids that are also present in immunoglobulin J segments and that are believed to be important for the structure of the immunoglobulin variable regions (see below). Interestingly, the T-cell receptor J gene segments are considerably more diverse in their 5' ends than the immunoglobulin J gene segments (Figure 6). This extra variability occurs in a portion of the molecule that for immunoglobulins falls within the third hypervariable region.

The murine J_β gene segments are grouped into two clusters, $J_{\beta 1}$ and $J_{\beta 2}$, each containing six functional gene segments and one nonfunctional gene segment (Figure 2). The J_β gene segments in each cluster are separated by 36-421 bp and are located 2-3 kb 5' to their respective C_β genes (71, 77, 78). Nucleotide sequence analysis of 27 J_α-containing cDNA clones has identified 22 different segments, implying that the repertoire of J_α gene segments is perhaps larger than 50 functional gene segments, and therefore much larger than that of the β locus which has 12 functional J gene segments, and the murine immunoglobulin gene loci which each have 3-4 functional J gene segments (94, 95). In addition, the organization of these gene segments is unique in that the mouse and human J_α gene segments are distributed over a large stretch of DNA (74-76). In mouse, the

J_α gene segments, span from 3 kb to greater than 63 kb 5' to the C_α gene (75) (Figure 2). Thus a V_α gene segment that rearranges to one of the J_α gene segments at the 5' end of the cluster, will be expressed on a very large primary transcript that would include a 63 kb intron. Preliminary data indicate the J_α gene segments are also further apart from one another than either the J_β, J_H or J_κ gene segments, as sequence analysis indicates that no two J_α gene segments studied to date are closer to one another than 500 base pairs (74-76). There are at least three murine J_γ gene segments; these are not organized in a J gene segment cluster, but instead each one is linked 3-4 kb 5' to a different C_γ gene (73). Therefore the J_γ-C_γ gene organization is similar to that of the immunoglobulin λ genes (Figure 2).

Constant Region Genes

C_β GENES In mice and humans there are two C_β genes, denoted $C_{\beta 1}$ and $C_{\beta 2}$ (77, 78, 96-99). The β locus is a tandem duplication of one D_β gene segment, a cluster of seven J_β gene segments and a C_β gene (Figure 2). The two C_β genes are very homologous to each other. In mouse the proteins encoded by these genes differ by four amino acids, in humans they differ by six, and in both species the differences are concentrated towards the 3' end of the gene (77, 78, 97, 99). In addition, the two human C_β genes share nucleotide sequence similarity extending into some of the introns (99). The similarity between the two mouse and human C_β genes could be due to either a recent duplication between these genes, selection for a conserved nucleotide as opposed to protein coding sequence, or gene conversion events. Because the similarity between the two C_β genes does not include all of the introns nor the 3' untranslated regions, the two C_β genes cannot have resulted from a recent gene duplication. Although selection for a nucleotide is possible, it is more likely that relatively recent gene conversion

events in each species are responsible for the similarity between the two C_β genes in each organism.

The murine and human C_β genes are encoded by four exons that, in contrast to the immunoglobulin C genes, do not correlate with the presumed functional domains of the constant region (77, 78, 99) (Figure 7). In the mouse, the first exon is 375 base pairs in length and encodes a block of 113 amino acids with homology to immunoglobulin constant regions followed by the first 12 amino acids of the connecting peptide. This is the only instance in which a block of sequence homologous to an immunoglobulin V or C region is not encoded in a separate exon in the immunoglobulin gene superfamily (see below). The 113 amino acid sequence contains two cysteines that could form a disulfide bond spanning 60 amino acids and is most homologous to the C_λ genes (37%) and the first domain of the immunoglobulin γ heavy chain constant regions (32-36%) (78). The remainder of the connecting peptide is encoded by a second exon of 18 nucleotides and a portion of the third exon of 107 base pairs. In immunoglobulin heavy chain constant regions the second exon encodes a proline-rich hinge region that permits flexibility in the constant region (100-105). Although the second exon of the C_β gene is positioned in approximately the same place as is the heavy chain hinge exon, it does not encode any prolines and thus does not resemble a hinge region. The remainder of the third exon encodes most of the transmembrane region, which consists of neutral and nonpolar amino acids, with the exception of a single lysine. The fourth exon of 179 base pairs encodes one amino acid of the transmembrane region, a 5-6 amino acid cytoplasmic region and the 3' nontranslated region.

C_α GENE There appears to be only one C_α gene in both mice and humans (37, 74-76). Like the C_β genes, the C_α gene consists of four exons (Figure 7). In the humans, the first exon is 261 bp long and encodes the region that is

homologous to immunoglobulin constant regions. However in the C_α gene this region is only 87 rather than 113 amino acids long and it contains a cystine bridge spanning only 49 amino acids, as opposed to the typical 60-75 amino acids. The second exon is 45 bp long and encodes 15 amino acids of the connecting peptide. The third exon, 108 bp, encodes the remainder of the connecting peptide, the transmembrane region, and a short cytoplasmic region, while the fourth exon encodes a 3' nontranslated region of 558 bp (76).

C_γ GENES There are three murine C_γ genes (73). One of these C_γ genes is a pseudogene in that a mutation in a donor splice recognition signal would probably prevent successful splicing of a C_γ transcript. In humans, there are two C_γ genes that are separated by approximately 6 kb of DNA (106) Unlike the C_α and C_β genes, the murine C_γ gene has only three exons (73) (Figure 7). The first exon, 330 bp long, encodes the region with homology to immunoglobulin constant regions; the second, 30 bp long, encodes a portion of the connecting peptide; the third, 545 bp long, encodes the remainder of the connecting peptide, the transmembrane sequence, a 12 amino acid cytoplasmic region, and the 3' nontranslated region.

EVOLUTION

Immunoglobulin Gene Superfamily

A gene superfamily is a set of multigene families and single copy genes related by sequence, implying a common ancestry, but not necessarily related in function (107, 108). The immunoglobulin and T-cell rearranging gene families are members of the immunoglobulin gene superfamily—named after the first well-studied members of this family (109). Over the past five years, gene cloning and nucleotide sequence analyses have lead to the identification of a number of other

genes that are members of this superfamily on account of their sequence similarity with immunoglobulins. Therefore the immunoglobulin gene superfamily now includes the three B-cell and the three T-cell rearranging gene families; the MHC-encoded class I heavy chain (110, 111) and class II (112, 113) genes; and a number of single copy genes including Thy-1 (108, 114) and MRC-OX2 (115), antigens of unknown function that are expressed in brain, in thymus and on other cell types; the poly-Ig receptor, which transports polymeric IgA and IgM immunoglobulin across mucous membranes (116); the T-cell accessory molecules T4 and T8 (116a, 117, 118); and β_2-microglobulin that is expressed in association with the heavy chain of the MHC class I molecule (119) (Figure 8).

Immunoglobulin Homology Unit

The polypeptide members of this gene superfamily are all constructed of one or more immunoglobulin homology units. Each homology unit is approximately 110 amino acids long and has several conserved amino acids and a centrally placed cystine disulfide bridge usually spanning 60-75 amino acids. From sequence analyses, it can be predicted that the immunoglobulin homology units all probably form a conserved tertiary structure denoted the antibody fold, which is composed of two sheets of three to four antiparallel β-pleated strands (120). Pairs of homology units can fold together in turn to create discrete polypeptide domains (e.g., V_L-V_H or C_L-C_H) characteristic of immunoglobulins. Thus the tertiary structure of one homology unit appears to facilitate interactions with a second homology unit to form a functional domain. At the DNA level, nearly every homology unit is encoded by a separate exon, demonstrating a correlation between the distinct structural features of these proteins and the exon-intron structure of the corresponding genes.

Genealogical Tree

A genealogic tree has been constructed for the immunoglobulin gene superfamily by assuming that evolutionary relatedness correlates with both the degree of sequence similarity among the members and with other features such as intron-exon structure and the ability to undergo DNA rearrangements (109) (Figure 8). The figure illustrates the hypothesis that many of the molecules important for the vertebrate immune response are encoded by genes that descended from a single ancestral sequence. According to this analysis, one of the earliest events in the evolution of the gene superfamily was a duplication leading to the divergence of V and C exons. Contemporary V and C homology units have little primary sequence similarity, suggesting an ancient divergence, although in immunoglobulins they retain similar tertiary structures. V and C homology units can be distinguished from one another by their length and by the presence of certain V- or C-specific conserved amino acids. Members of the immunoglobulin gene superfamily are composed of different numbers of either V and/or C homology units. A second critical early event was the acquisition of the ability to rearrange DNA, which may have arisen from the capture of a complex transposon by a primordial V gene (121).

REARRANGEMENT AND EXPRESSION OF α, β AND γ GENES

A large panel of lymphoid cells have been tested for rearrangement and expression of T-cell receptor α, β and γ genes. Several questions have been addressed by these analyses. First, in light of the dual specificity of T lymphocytes for antigen and self-MHC-encoded molecules, is it possible that two T-cell antigen receptor molecules are expressed on individual T cells? This could occur if other receptor gene families besides α and β are expressed, or if the expression of the V genes in at least one of the defined receptor gene families is

not subject to allelic exclusion. Second, are there antigen receptor isotypes that are associated with the different functional categories of T lymphocytes? Finally, are gene segment rearrangements regulated so that they occur in a well-defined progression, and, if so, what insight does it give us into the ontogeny of T lymphocytes and into possible mechanisms leading to allelic exclusion?

Tissue-specific DNA Rearrangements

The available data summarized above indicate that the rearrangement processes in the immunoglobulin and T-cell specific gene families are quite similar. As mentioned above, similar recognition sequences for DNA rearrangement are used in all six gene families (Figure 3), and the one-turn to two-turn rule of the recognition sequences for gene segment joining is always followed. In addition, the rearrangement processes in B and T cells are imprecise, resulting in the addition and/or deletion of nucleotides at the junction of the joined gene segments (see below). In spite of these similarities, the complete rearrangement of T-cell antigen receptor and immunoglobulin genes appear to be tissue-specific. Approximately 10% of mouse T lymphocytes have D_H-J_H rearrangements (122-126), but V_H (127-130) and light chain gene segments do not rearrange in these cells (122, 131). Similarly, only six cell lines out of a sample of more than 100 B lymphocyte tumors and hybridomas that were examined have undergone β gene rearrangement (F. Alt, M. Kronenberg, R. Perlmutter, unpublished observations and refs. 132, 133, 134) and there is no evidence to indicate that these events lead to the production of a functional β polypeptide. Insight into this tissue-specific regulation has emerged from studies in which DNA constructs containing unjoined D_β and J_β gene segments were introduced into B-cell lines that carry out immunoglobulin gene segment rearrangements (G. Yancopoulos, K. Blackwell, L. Hood & F. Alt, unpublished data). In these cells, the β gene segments are rearranged in a site-specific fashion as efficiently as D_H and J_H gene segments

introduced in a similar manner. These data suggest that the DNA recombinational machinery in B and T cells is very similar and that other cell type-specific features such as chromatin structure may determine which gene families rearrange in the two cell types.

Mechanisms of DNA Rearrangement

Several mechanisms have been proposed to account for immunoglobulin gene segment rearrangements (Figure 9). The DNA between gene segments may loop out, the stem of this loop being formed by base pairing between the recognition signals for DNA rearrangement (Figure 3a). The stem-loop structure may then be excised to complete gene segment joining. Most of the rearrangements in the immunoglobulin heavy chain gene family result in the deletion of the DNA sequences between the joined gene segments and are therefore consistent with the deletion model (135). However, κ gene rearrangements often appear incompatible with this deletion model (136-141) and consequently three other models for rearrangement have been proposed: homologous but unequal sister chromatid exchange (137, 138, 140), inversion (139, 142, 143), and reintegration of the deleted sequences (136, 137) (Figure 9).

Several of these models are required to explain the rearrangements that occur in the β gene family. Although the majority of T cells (24/37) that were analyzed in detail had a β gene rearrangement pattern that is consistent with the deletion model, 13 out of 37 T cells had Southern blot patterns that could not be explained in this way (144). In some cases, the DNA in the region between joined gene segments was retained in the genome, an observation consistent with any one of the three additional models (144). In other cases, a partial duplication of the J_β-C_β locus was observed, an observation consistent only with homologous but unequal sister chromatid exchange (144, 145).

The detailed analysis of all the β gene segment rearrangements in one T_H

cell line has shown that the V_β gene segment expressed in this cell had rearranged by inversion (86). As noted above, in germline DNA one of the V_β gene segments, $V_{\beta 14}$, is located 10 kb 3' to the $C_{\beta 2}$ gene and in the opposite transcriptional orientation to the J_β gene segments and C_β genes (Figure 2). Two different mechanisms were required to form the complete V_β gene in this cell (Figure 10). First, the $D_{\beta 1}$ gene segment joined to the $J_{\beta 2.3}$ gene segment via a deletion or homologous but unequal sister-chromatid exchange. The $V_{\beta 14}$ gene segment then rearranged, inverting a 15 kb sequence of DNA. One inversion breakpoint contains the joined V_β-D_β-J_β gene, while the other encompasses the reciprocal recombination product containing the heptamer 3' of the V_β gene segment joined to the heptamer 5' of the $D_{\beta 1}$ gene segment (86). Thus, taking into account all the data, it appears that all three of the mechanisms depicted in Figure 9 may be employed for β gene rearrangements.

The detailed mechanisms of rearrangement in the α and γ gene families have not been so well characterized. It is also not clear why rearrangements in some of these gene families are always characterized by deletions, as observed for the immunoglobulin heavy chain genes, while in others, such as the κ and β gene families, different mechanisms also appear to operate. Any V gene segment such as $V_{\beta 14}$ that is in the opposite orientation to the J gene segments, must undergo an inversion in order to be expressed. Perhaps an inversion of germline DNA that included some V gene segments occurred during the evolution of both κ and β gene families, therefore requiring a second inversion as a mechanism for their somatic rearrangement and expression. The factors that might facilitate unequal sister chromatid exchanges in some but not other gene families, however, are unclear.

Allelic Exclusion

The rearrangement of β genes in T cells normally occurs on both chromosomal homologues. In a panel of 37 clonal T lymphocyte lines, only 3% of

the $J_{\beta 1}$ gene segment clusters and 13% of the $J_{\beta 2}$ gene segment clusters were in a germline configuration (144). Densitometric analysis of Southern blots from heterogeneous T-cell populations demonstrates that normal T cells also display a similar degree of β gene rearrangement (146, 147).

Southern blot analysis of three T cell lines and the isolation of all the β gene rearrangements from three others, indicates that each of these six T cells can express only a single V_β gene (86, 144, 148). Therefore, although β gene rearrangement occurs on both chromosomal homologues, the expression of these genes may be allelically excluded in that only one allele is productively rearranged and used in the synthesis of a functional β chain in individual T cells. Whether a β gene rearrangement is productive is determined by the joining of the V_β gene segment, which sets the translational reading frame of the J_β gene segments. Since only one of the three possible translational reading frames is productive for each J_β gene segment, and because of the apparently random addition and deletion of nucleotides at the V_β-D_β junction, in two-thirds of the V_β-D_β-J_β rearrangements the J gene segments will not be in the proper translational reading frame.

The α and γ gene rearrangements generate junctional and possibly N-region diversity as well (73-75, 149) (see below). As a consequence, many rearranged genes in these families should not be in the proper translational reading frame. There is not yet enough information to determine whether α and γ gene rearrangements are allelically excluded, although this seems likely. Southern blots of four cytotoxic T-cell clones initially analyzed showed they had undergone γ chain gene rearrangements on both homologues (31), and one of these T-cell clones that was analyzed in more detail had one productive and one nonproductive γ gene rearrangement (149).

Beta Gene Rearrangements in Different Cell Types

HELPER AND CYTOTOXIC T CELLS AND T-CELL TUMORS β gene rearrangement is nearly ubiquitous in human and murine T cells. In humans, 4/4 helper T-cell lines, 7/7 cytotoxic T-cell lines, one cell line with both cytotoxic and helper function (150, 151), and 65/71 T-cell tumors exhibited β gene rearrangement (132-134, 152). The remaining six tumors had germline β genes and they may be similar to immature T cells, or they may not in fact belong to the T-cell lineage. Similarly, in the mouse 28/28 helper T-cell lines and hybridomas, 16/16 cytotoxic T-cell lines, and 14/14 T-cell tumors exhibited β gene rearrangements (144, 153). Therefore all cytotoxic and helper cells rearrange and probably express β chains, although in many cases there is no conclusive evidence that these rearrangements are productive.

SUPPRESSOR T-CELLS Suppressor T cells are defined by a number of different and complex immunological assays. Of 15 mouse suppressor T-cell hybridomas tested, two exhibited β gene rearrangements, one had germline β genes, and the remaining 12 hybrid cells had apparently deleted the β gene loci from both chromosomes that were contributed by the suppressor T-cell fusion partner (144, 153). Therefore, the T_S cells in these cases do not utilize the β chain in their antigen receptors. In contrast, five of five human suppressor T-cell lines tested exhibit β gene rearrangements and there is evidence indicating that some of these cells express a typical α/β heterodimer (150, 151, 154). These data imply that there may be several classes of T_S cells, only some of which utilize the β genes, although other explanations are possible.

NATURAL KILLER CELLS Lymphoid cells that spontaneously exhibit cytotoxic activity against a variety of tumor and nontransformed cell types, are known as natural killer (NK) cells. NK cells can have several β gene phenotypes. Cloned murine NK lines that are cultured in the presence of IL-2 have β gene rearrange-

ments and transcripts (155), although several rat large granular lymphomas that display NK activity have germline β genes (156). Peripheral blood cells from normal human donors that express cell-surface markers characteristic of NK cells also had β gene rearrangements, although there was significantly less rearrangement than was observed in T-cell populations (147). This result is consistent with the results of an analysis of a panel of human clones with NK activity, those that express T3 have β gene rearrangements and α transcripts, while those that do not express T3 have no α transcripts and may only have D_β-J_β rearrangements (157). Collectively these results indicate that NK cells are a mixed population with regard to β gene rearrangement, and that expression of an α/β heterodimer is not required for NK activity. It remains possible, however, that in some cases T-cell antigen receptors participate in the recognition of NK target molecules.

Transcription of Beta Genes

There are two predominant size classes for the β transcript: a 1.3 kb RNA that contains a V_β gene and a 1.0 kb RNA that does not (82, 144, 150). The 1.0 kb transcripts appear to be derived from rearrangements that joined a D_β to a J_β gene segment in the absence of a V_β gene segment rearrangement (82). In the mouse, both $D_{\beta 1}$ and $D_{\beta 2}$ gene segments have been identified in cDNA clones that contain D_β joined to J_β gene segments but that lack V_β gene segments, implying that promoters exist in the 5' flanking regions of both germline D_β gene segments (81, 82). Because the transcribed 5' flanking regions also contain open reading frames and an in-frame methionine codon that could serve as a start-signal for translation (82), the sequences of these rearranged D_β-J_β genes could encode truncated polypeptides containing sequences 5' to the D_β gene segments, as well as D_β, J_β and C_β sequences. Although there is so far no evidence for these truncated β polypeptides, in immunoglobulins the analogous D_H-J_H rearrangements do give rise to both transcripts and truncated polypeptides containing the

C_μ gene product (158). Other transcripts that may be derived from unrearranged genes and several aberrantly spliced β transcripts have also been observed (R. Barth, unpublished observations and ref. 159).

Alpha Gene Rearrangement and Transcription

The C_α gene is transcribed in a variety of cell types including helpers, cytotoxic cells, many tumors, and the T3 positive NK clones mentioned above. Two size classes of α chain RNA have been described, a 1.7 kb transcript that contains V_α sequences and a 1.4 kb transcript (32, 33). By analogy with the β chain genes, the shorter transcript could be derived from a D_α-J_α rearrangement or from a J_α-C_α transcript initiated by a promoter 5' to the J_α gene segment. Consistent with the latter hypothesis is the identification of a cDNA encoding a germline J_α-C_α transcript (A. Winoto, unpublished). The levels of α and β RNA in peripheral T cells are roughly equivalent (32, 146, 160). However, the thymus contains far more β than α RNA, consistent with the hypothesis that the β genes rearrange and are expressed before the α genes (32, 146) (see below).

Gamma Gene Rearrangement and Transcription

Transcripts containing γ gene sequences were found in 4/4 cytotoxic cells tested, but in only 1/10 helper T-cell lines and hybridomas (31, 149) and none of the human NK clones tested (157). This pattern of expression suggests that the γ chain and MHC class I recognition or antigen-specific T-cell cytotoxicity are in some way linked. Compared to β transcripts, γ RNA is at least 10-fold less abundant in both adult thymocytes and peripheral T cells (32, 146, 160).

While transcription of γ gene is confined primarily to cytotoxic cells, γ gene rearrangement is equally prevalent in both helper and cytotoxic T-cell populations. DNA prepared from Lyt-2 positive cells that are predominantly cytotoxic and suppressor T cells, and DNA prepared from L3T4 positive cells that

are predominantly helper T-cells, had approximately the same amount of γ gene rearrangement (149). In addition, 8/10 IL-2 producing T-hybridomas restricted by class II MHC molecules have γ gene rearrangements and no germline γ chromosomes (R. Haars, J. Kobori, N. Shastri, unpublished data and ref. 149). The other two hybrid cells deleted the γ genes that were not derived from the BW5147 fusion partner. All of the $T4^+$, $T8^-$ human cell-lines tested also had γ gene rearrangement (106). It is not yet known whether any of these rearrangements are productive.

All or most of the γ gene rearrangements in murine T cells are similar, suggesting that murine V_γ and J_γ gene segment diversity is quite limited (146, 149). In the human, most cell lines have different rearrangements, suggesting that V_γ gene segments may be more diverse (106).

T-Cell Receptor Isotypes

Isotypes are defined as multiple nonallelic forms of constant region genes. The immunoglobulin heavy chain gene locus is characterized by several C region isotypes that participate in different effector functions (reviewed in 161). There is no evidence for C_α isotypes; only one C_α gene hybridizes with the available probes (37, 74, 75), and this C_α is transcribed in a variety of cell types (32, 33). The $C_{\beta 1}$ and $C_{\beta 2}$ genes are isotypes, yet there is no evidence that expression of either one of these two C_β genes is related to the ontological development of a T cell (146, 160), its function or the specificity for either antigen or MHC molecule. For example, both helper and cytotoxic T cells can express either the $C_{\beta 1}$ or the $C_{\beta 2}$ genes (144). The $C_{\beta 2}$ transcript is more abundant than $C_{\beta 1}$ in RNA from lymphoid tissues (146), and T-cell clones use the $C_{\beta 2}$ gene more often than the $C_{\beta 1}$ gene (88, 144). However, this bias probably reflects an inherent statistical bias for rearrangement of D_β gene segments to the $J_{\beta 2}$ gene cluster (see below). Finally, the functional equivalence of $C_{\beta 1}$ and $C_{\beta 2}$ is best

demonstrated by NZW mice that lack a $C_{\beta 1}$ gene, as well as $D_{\beta 2}$ and $J_{\beta 2}$ gene segments, and have apparently normal T-cell function (162). Since it appears that all helper and cytotoxic and some suppressor T cells employ an α/β heterodimer, the T-cell antigen receptor apparently lacks the equivalents of both functional heavy chain isotypes and the immunoglobulin heavy-chain class switch.

Unlike expression of the α and β genes, the pattern of γ gene transcription suggests that expression of this molecule, possibly in association with some other cell-surface protein, is related either to class I MHC recognition or cytotoxic function. Analysis of γ gene expression in MHC class II-specific cytotoxic T cells should establish whether the type of MHC molecule recognized or T-cell function correlates better with expression of this molecule. The lack of γ gene transcripts in some cell types, however, raises the possibility of an isotypic γ chain functional equivalent that could be expressed by these cells.

THE ONTOGENY OF T-CELL ANTIGEN RECEPTOR REARRANGEMENT AND EXPRESSION

The ontogeny of α, β and γ gene rearrangement and expression in fetal and adult cells has been studied by several groups. Two major issues have been addressed by these studies. First, do rearrangements occur in a sequence of steps as is the case in the three immunoglobulin gene families? If rearrangements occur in this manner it might be possible to unambiguously assess the developmental stage of some T cells by determining which gene families had and which had not rearranged. Second, when and where in development are these genes first expressed?

Ordered Rearrangement and Expression

ORDERED EXPRESSION OF IMMUNOGLOBULINS The study of the

developmental control of T-cell antigen receptor and γ gene rearrangement and expression has been influenced by studies on the ontogeny of immunoglobulin rearrangement and expression in B cells. Based on these immunoglobulin gene studies, a regulated model for the rearrangement of gene segments in B cells has been proposed that accounts for the observed rearrangements in developing B cells and proposes a mechanism for allelic exclusion. Immunoglobulin rearrangement and expression can be divided into four separate stages (reviewed in 163). The first stage, D_H-J_H rearrangement, occurs in pre-B cells and is believed to be independent of allelic exclusion (135, 164). The second stage involves the rearrangement of the V_H gene segment to the joined D_H-J_H gene segments. According to the regulated model, if this second rearrangement event is non-productive, further gene segment rearrangements occur on the other chromosomal homologue (135). A productive V_H-D_H-J_H rearrangement is believed to inhibit further heavy chain gene rearrangements; it has been proposed that the synthesis of a heavy chain polypeptide plays a role in this process (165). The third step involves the rearrangement of the V_κ and J_κ gene segments (166). Because this appears to require productive heavy chain gene rearrangement, the heavy chain polypeptide may also play a role in activating gene rearrangement in this family. The successful rearrangement and expression of a κ protein and the subsequent formation of a complete cell-surface immunoglobulin molecule is believed to be involved in the termination of further κ rearrangements (167, 168). The fourth step is the rearrangement of the λ light chain genes, which only occurs if the κ genes have rearranged nonproductively on both homologues (167, 169).

FETAL TISSUES AND T CELLS To determine if there is an order to the rearrangement and transcription of the α, β, and γ genes, RNA and DNA from the appropriate fetal tissues have been hybridized with DNA probes. The results of these studies are summarized in Table 2. The mouse has a gestation period of

approximately 20 days, and immunocompetent thymocytes can be detected shortly before birth (170). The thymus itself develops from nonlymphoid cells in the pharyngeal pouch and cleft, and other tissues, and is first seeded by lymphocytes around the eleventh day of gestation (171). Most of the prethymic cells originate in the fetal liver. The thymus increases dramatically in cell number between day 12 and day 18, due to both cell migration from fetal liver and proliferation of thymocytes.

Analysis of the transcripts present in fetal tissues indicates that γ RNA appears first, followed by β and finally α, and that the level of γ RNA decreases around the time that the α message appears (146, 160, 172, 173) (Table 2). Although no transcripts were detected in fetal liver from any stage, γ RNA could be detected in day 14 fetal thymus, the earliest day tested (146). At this stage, the thymus contains mostly developing T cells. The Lyt-2 and L3T4 molecules are not yet expressed (174-176), but most day 14 fetal thymocytes express the Thy-1 antigen (175, 176), and some express IL-2 receptors (177, 178). No C_α RNA can be detected at this stage (146, 160, 173). A 1.0 kb transcript derived from D_β-J_β rearrangements is present, but 1.3 kb transcripts containing V_β genes can first be barely detected on day 15 (146, 172). On day 16, a relatively large amount of 1.3 kb β gene RNA is present (146, 160, 172) and α gene transcripts can be detected, although most of the α RNA is found in the shorter 1.4 kb RNA transcript (146). Substantial levels of both α and β RNA are present by day 17 (146, 160, 172, 173). The presence of significant amounts of complete α and β gene transcripts on day 17 is consistent with biochemical and immunochemical studies demonstrating that cell-surface expression of α/β heterodimers can first be detected on this day (172, 179).

Analysis of DNA prepared from fetal organs or from hybridomas made by fusing fetal cells with the BW5147 T lymphoma indicates that the order of re-

arrangement of the β and γ genes in fetal tissues is consistent with the order of expression: namely, V_γ joins to a J_γ gene segment, around the same time or soon afterward D_β-J_β gene segment joining occurs, followed by V_β rearrangement (146, 180).

IMMATURE ADULT THYMOCYTES Experiments on selected populations of adult thymocytes have yielded results similar to those obtained with fetal thymocytes of day 14-16 gestation. The Lyt- 2^-, L3T4$^-$ or double negative subpopulation resembles fetal thymocytes from days 14-15 gestation in that both populations do not express the Lyt-2 and L3T4 antigens, both are rapidly dividing, and both probably rearrange and express the α, β, and γ genes in a similar order (160, 181, 182) (Table 2). These double-negative cells, which are thought to give rise to the other thymus subpopulations (reviewed in 183, 184), have large amounts of γ (160) and very little, if any, α RNA (181, 185). In addition, although most of the cells in the adult thymus have already undergone β gene rearrangements on both homologues, some of the double-negative cells have germline β genes (181, 182). A subpopulation of the double negative cells expresses the cell-surface antigen Pgp-1, which is also found on bone marrow cells and on many day 14-15 fetal thymocytes. This subpopulation, which may represent the earliest lineage in the adult thymus, has almost no β gene rearrangement (182). Therefore, despite the fact that the different immature subpopulations are a minute fraction of the adult thymus and are difficult to separate from one another, the γ-β-α progression also appears to occur in these cells.

In summary, although it is still controversial as to whether a single differentiation pathway characterizes T-cell development, or whether cortical and medullary cells differentiate separately (11, 12, 184, 186), the results are consistent with a single ordered pathway for the rearrangement of γ, then β and then α genes in all T cells (Figure 11). In the proposed pathway, the V_γ genes

rearrange and are transcribed first, followed by rearrangement and transcription of joined D_β-J_β gene segments, followed by rearrangement and transcription of V_β gene segments and then finally by rearrangement and transcription of V_α gene segments. If either D_α or D_γ gene segments exists, then further discrete steps may be identified. The proposed pathway is well supported by the analysis of whole organs, cell populations and a large number of hybridomas; however, there is so far no T-cell equivalent of the Abelson-transformed B-cell lines and, therefore, the differentiation steps that a single cells undergo cannot be directly followed.

The apparent regulation of gene rearrangement in T cells is consistent with the model described for immunoglobulin genes. As outlined above, according to this model rearrangement proceeds in a stepwise fashion, the result of a particular step is tested before proceeding to the next step. Applying this model to T cells, a productive V_β gene segment rearrangement may shut down further V_β gene segment rearrangement through the synthesis of a β polypeptide, while a nonproductive rearrangement does not. Applying this model further, it is possible that a productive β gene rearrangement is a prerequisite for α gene rearrangement. However, γ gene rearrangment may be regulated differently, because although this gene family rearranges early, data from helper cells suggests that productive γ gene rearrangement may not be required for further T-helper cell development. We might therefore speculate that productive γ gene rearrangement is a developmental switch-point, in that cells with productive γ rearrangements tend to become either MHC class I restricted or cytotoxic cells, while cells with only nonproductive rearrangements tend to become MHC class II restricted or helper cells (Figure 11).

The Site of Rearrangement and Expression

INTRATHYMIC AND EXTRATHYMIC REARRANGEMENTS Several

experiments strongly suggest that thymic precursors do not express T-cell antigen receptors and that α and β gene segment rearrangement and expression occur following migration to the thymus, although some γ gene rearrangement may occur prior to migration of precursors to the thymus.

Two results demonstrate that some γ gene rearrangement occurs prior to entry into the thymus. First, a rearranged band can be detected in DNA prepared from whole day 13 fetal liver, but not in fetal liver from later stages of development (146). To be detected in a heterogeneous cell population, this rearrangement must have occurred on at least 5% of the chromosomes in the fetal liver cells. The same rearranged fragment is also found in day 14 and day 15 fetal thymocytes, suggesting a migration of cells with this rearrangement from the liver to the thymus on days 13-14 (146). Second, fetal liver x BW5147 hybridomas made from cells at various stages of gestation have γ gene but not β gene rearrangement. The discrepancy between the lack of γ RNA in fetal liver and the observed rearrangements in a few fetal liver cells may be due to the very low levels of this transcript.

The results of several experiments suggest that expression of α and β chains is initiated by gene rearrangement events that occur in the thymus. First, complete α and β transcripts cannot be detected in fetal liver, and they are not present in the most immature adult thymocytes or in fetal thymus until the 17th day of gestation (146, 160, 173). Second, there appears to be a gradient of increasing β gene rearrangement in thymocytes as development progresses (146, 180); most of the cells have germline β genes on day 14, D_β-J_β gene segment rearrangements appear soon afterwards, and by day 17 most of the chromosomes have either D_β-J_β, or V_β-D_β-J_β rearrangements (146, 180). It is unlikely that these changes in the thymus are caused by seeding of this organ by waves of increasingly mature precursors from the fetal liver as cultures of day 14 and 15

thymus tissue have demonstrated that these cells can differentiate and attain immunocompetence *in vitro* (174, 175, 187-192). Some nonproductive β gene rearrangement in day 13 fetal liver has been observed (146), and the possibility of some productive prethymic rearrangement of α and β genes cannot be formally ruled out. However, all the data are consistent with the hypothesis that thymic precursors do not express α and β polypeptides, and that productive rearrangement and expression of these genes takes place in the thymus.

NONPRODUCTIVE REARRANGEMENTS AND THYMIC T-CELL DEATH If the α and β genes rearrange in the thymus, and cells that have only nonproductive α or β gene rearrangements do not emigrate, then a large part of the intrathymic cell death that occurs in the adult thymus can be attributed to these nonproductive rearrangements. As mentioned above, two out of every three V_β gene segment rearrangements will be nonproductive, as they would place the J_β gene segment in an inappropriate translational reading frame. This would mean that 0.67 x 0.67 = 0.44, or 44% of the cells will have nonproductive β gene rearrangements on both homologues and therefore may not survive to emigrate from the thymus; the remaining 56% might go on to rearrange their α genes. If we assume either a direct V_α-J_α joining or a V_α-D_α-J_α mechanism similar to what is observed in β genes, then two of three V_α rearrangements will also be nonproductive. Again, 44% of the remaining T cells will have only nonproductive α gene rearrangements and conversely, 56% of the remaining T cells will have a productive rearrangement. If we consider the progeny of a pre-T cell undergoing gene segment rearrangement in the thymus, no more than 0.56 x 0.56 = 0.31, or 31% of the cells will produce a functional α-β heterodimer. Thus, a large percentage of the intrathymic cell death could be due to the lack of successful rearrangements in the α and/or β genes. This hypothesis predicts that many of the cells that are destined to die in the thymus should not express T-cell antigen

receptors on their surface. Recent experiments that have addressed this issue are not entirely conclusive. While it is certain that many of the cells with a surface phenotype thought to be characteristic of those that die in the thymus have α and β transcripts (185), there are conflicting estimates of the fraction of cells in this category that express α/β heterodimers (174, 193, 194).

GENERATION OF DIVERSITY IN THE VARIABLE REGION GENES

The problem of generating diversity in the lymphocyte antigen receptor is common to both B and T cells. The immunoglobulin genes utilize three basic mechanisms: germline diversity, the utilization of a large number of different gene segments for variable region formation; combinatorial joining, the random rearrangement of the different germline gene segments; and somatic mutational mechanisms that occur during and after the formation of the variable region gene. A number of experiments have been conducted to determine to what extent these mechanisms are utilized in the T-cell antigen receptor.

Germline Diversity

BETA GENES The mouse germline β gene repertoire is composed of a relatively limited number of V_β and D_β gene segments and a relatively large number of J_β gene segments. Characterization of expressed V_β genes from murine thymus, spleen, T-cell hybridomas, and functional T-cell lines has revealed that a limited set of V_β gene segments is utilized in these T cells. Of 37 that were sequenced, only 16 different V_β gene segments have been found (87-89, 148, 159). Assuming that the V_β gene segments are expressed with equal frequency, the total number of functional germline V_β gene segments can be expected to be 24 or less at the 95% confidence level (88, M. Meister, unpublished). Although the numerical calculation may be unreliable because the different V_β gene segments

are not expressed with equal frequency (87), only a few V_β gene segments are expressed by many different clones in several different mouse strains, which is unlikely if there were a large V_β gene segment repertoire. Southern blot analyses using the different V_β gene segment probes also support the notion of a limited repertoire (87-89). Surprisingly, several mouse strains have deleted a significant fraction of this repertoire and can still survive in the laboratory. In SJL mice, the best studied example, 10 V_β gene segments from six subfamilies have been deleted from the germline repertoire (R. Barth and B. Kim, unpublished observations and D. Loh, personal communication, 89), and there is no evidence so far in favor of extra V_β gene segments compensating for this deficiency (D. Loh, personal communication). However, studies on human β gene cDNA clones have indicated that the human V_β gene segment repertoire is likely to be much larger than that of the mouse (P. Concannon, L. Pickering & L. Hood, unpublished).

Allowing for diversity arising from the recombination process, all of the D_β regions in murine V_β genes could be encoded by one of the two known germline D_β gene segments (87-89) (Figure 11). However, in some cases only a few central nucleotides are shared with these D_β gene segments, and the existence of other germline D_β gene segments cannot be ruled out. Preliminary sequence data indicate that, like the human V_β repertoire, the human D_β gene segment repertoire may be larger than that of the mouse (P. Concannon, L. Pickering & L. Hood, unpublished). Of the 14 murine J_β gene segments in the two J_β clusters, one in each cluster is a pseudogene and the other 12 are apparently functional (71, 77, 78). Ten of these 12 J_β gene segments have been found in the V_β regions that have been sequenced so far.

ALPHA GENES The germline V_α and J_α gene segment repertoires are larger than the corresponding β gene segment repertoires. A number of murine V_α gene segments obtained from cDNA libraries constructed from thymus and

functional T-cell hybridomas have been sequenced (32, 33, 94, 95). The 25 different cDNA clones analyzed have 22 different V_α sequences that can be grouped into 11 subfamilies. Southern blot analyses with V_α probes from nine different subfamilies reveal that there are at least 55 bands that cross-hybridize with these probes (32, 33, 94, 95, and J. Klotz, unpublished), indicating that the germline V_α gene segment repertoire is larger than the V_β gene segment repertoire in mouse. The identification of 11 different V_α gene segment subfamilies is more than that observed so far in the murine immunoglobulin families and comparable to that of the β gene family. As noted previously, the J_α gene segment repertoire is much larger than the J gene segment repertoire in any of the other B- or T-cell receptor gene families (94, 95).

GAMMA GENES There are three V_γ gene segments that cross-hybridize with the V gene segment present in the original murine γ gene cDNA clone and there are three murine J_γ gene segments (74). Although joining of these V and J gene segments permits nine different rearrangements, only three rearrangements account for most of those observed in murine T cells (146, 149). One of these prevalently rearranged γ gene restriction fragments does not hybridize to the V_γ probe (149) and it therefore could involve another V_γ gene segment, or alternatively, it could result from a D_γ-J_γ rearrangement. The remaining two non-germline fragments hybridize with the V_γ probe (149). However, one of these rearrangements joins a V_γ to a J_γ gene segment adjacent to the C_γ pseudogene (73) and accordingly is not functional. Therefore it is possible that of the three predominant γ rearrangements, only one involving a particular V_γ and J_γ gene segment leads to the expression of a γ polypeptide. It is not clear why the other two known V_γ gene segments are rearranged infrequently.

Combinatorial Joining

Analysis of expressed murine V_β genes and β gene rearrangements in T-cell

hybrids and clones indicates that combinatorial joining occurs in β genes (87-89). For example, the $D_{\beta 1}$ gene segment joins to J gene segments in either cluster with approximately equal frequency (88, 180). In contrast, if we consider only rearrangement by deletion, the $D_{\beta 2}$ gene segment can only join to $J_{\beta 2}$ gene segments (87-89). The available sequences are also consistent with the hypothesis that V_β gene segments may join to any D_β-J_β rearrangements (87-89). If these V_β rearrangements are equally likely, there should be a bias in favor of $J_{\beta 2}$ and hence $C_{\beta 2}$ expression, because one germline D_β gene segment can join to either $J_{\beta 1}$ or $J_{\beta 2}$ gene segments while the other can only join to $J_{\beta 2}$ gene segments. Such a bias in favor of $C_{\beta 2}$ expression is in fact seen in the analysis of different cDNA clones or of RNA from lymphoid tissues. There is also evidence supporting combinatorial joining of V_α and J_α gene segments (A. Winoto, unpublished). As noted above, the extent of combinatorial joining in the γ genes may be quite limited.

In addition to the random usage of gene segments, the β genes have an additional feature that may permit either optional or multiple usage of the D_β gene segments. The V_β gene segments have a two-turn recognition signal for DNA rearrangement (71, 72), the D_β gene segments have a one-turn recognition signal in the 5' and a two-turn recognition signal in the 3' flanking region (80-82), and the J_β gene segments have a one-turn recognition signal (71, 72, 77, 78) (Figure 3). In principle the V_β gene segment can therefore rearrange directly to a J_β gene segment and still obey the one-turn to two-turn joining rule. Similarly, D_β-D_β joinings could be possible to create longer D_β regions. Either of these events would increase the diversity of the junctional region of the V_β gene. Analyses of different V-D-J junctions have identified a few potential V-J and D-D joining events (97, 98, 159) although there are other interpretations for these examples. If direct V_β-J_β joining or D_β-D_β rearrangements do occur, it is clear that these events are infrequent.

Somatic Mutational Mechanisms

The α, β and γ genes employ three different mechanisms for somatic mutation that occur during the formation of the V gene. These mechanisms also occur in the immunoglobulin V gene families, although the use of a fourth mechanism found in immunoglobulins is doubtful for the T-cell receptor.

JUNCTIONAL DIVERSITY The process that joins two gene segments together is imprecise. The joining event can delete nucleotides from the ends of the V, D and J gene segments, leading to codon changes at the junctions of these segments (Figure 12). This process, called junctional diversity, occurs in both the immunoglobulin (123, 195, 196) and T-cell receptor (73, 75, 76, 81, 82, 148, 149) gene families.

N-REGION DIVERSITY Additional nucleotides not encoded by either gene segment are added at the junction between the joined gene segments during rearrangement (Figure 12). This mechanism, known as N-region diversification, is utilized by the immunoglobulin V_H genes, but not by the V_κ or V_λ genes (195, 197). N regions are also clearly present in the T-cell V_β genes (81, 82, 148); the occurrence of N-region diversification in these four gene families is therefore correlated with the presence of D gene segments. However, this correlation may not include all six rearranging gene families because extra nucleotides are also seen in the V-J junctions of α and γ genes (73-76). It is not known whether these α and γ sequences arise from either a combination of N-region diversification and D gene segments, D gene segments alone, or from N-region diversification in the absence of D gene segments.

MULTIPLE TRANSLATIONAL READING FRAMES IN THE D GENE SEGMENT Analysis of the D_β regions of different V_β genes has shown that the two germline D_β gene segments are utilized with approximately equal frequency in all three translational reading frames (88, 148) (Figure 13). Although this adds

to the diversity of V_β gene segments, the low sequence complexity of the two germline D_β gene segments limits the contribution of this mechanism. In immunoglobulin V_H genes, however, the D_H gene segments are used in the same reading frame in most of the cases analyzed. The statistical preference for a single reading frame for the D_H gene segments cannot be accounted for simply by the presence of a stop codon in one reading frame of some of the D_H gene segments. The ability to use the D_β gene segments in three reading frames need not be considered an actual mechanism, but may instead reflect the fact that there is little selection for particular protein sequences in this part of the molecule. Despite this coding sequence flexibility, the germline nucleotide sequences of mouse and human $D_{\beta 1}$ are identical, suggesting that there may be some selective pressure for the maintenance of a particular germline DNA sequence in this region of the β locus.

SOMATIC HYPERMUTATION Immunoglobulins employ an additional mechanism for diversification, somatic hypermutation, that generates base-pair substitutions throughout the V gene and its flanking sequences at a late stage in B-cell development (198-200). This mechanism is capable of generating up to 3% sequence differences between the mutant and the corresponding germline gene segment (199). Those B-cells with receptors that have higher affinities for antigen as a result of somatic hypermutation are believed to be selectively expanded when the antigen concentration is limiting (201).

Analyses of several different expressed V_β genes have shown that with the exception of the differences in the junctional regions, they are identical in sequence to their germline gene segment counterparts (71, 93, 148). Substantially more data is available if the nucleotide sequences of different cDNA clones are compared to one another, including several cases where no germline sequence is available. T cells expressing the same germline gene segments most often have

identical gene segment sequences (71, 87-89, 159). In those cases where the sequences are not identical, there are very few differences in nucleotide sequence and never more than two substitutions at the amino acid level. The observed differences could be due to either somatic hypermutation, interstrain sequence polymorphisms or they could be due to nucleotide sequencing errors (87-89). The preliminary nucleotide sequence data of V_α genes suggests that somatic hypermutation is also rare in this gene family. Two different sets of identical, expressed V_α gene segments have been found: the first from a collection of thymus cDNA clones (94), and the second from a series of cytochrome c/I-E^k-specific T hybridomas (A. Winoto, unpublished results). Because in each case the three identical V_α gene segments were associated either with different J_α gene segments, or the V_α-J_α junctional region was different, the independent origin of each of these clones is established. Taken together, these data indicate that somatic hypermutation occurs infrequently, if at all, in the T-cell receptor genes.

There has, however, been a report of somatic variants arising in alloreactive T hybridomas in which new specificities were generated against class II molecules after more than 200 cell generations *in vitro* (202). Because of the small number of mutant cells isolated, the frequency with which similar events occur has not been established. Therefore it is not clear that the mechanism by which these mutants arise is comparable to that producing hypermutation in immunoglobulins, and the physiologic relevance of this provocative finding is uncertain.

Extent of T-Cell Receptor V, D and J Diversity

Although the T-cell antigen receptor utilizes the same mechanisms for generating diversity as the immunoglobulin genes, there are several differences in the extent to which these mechanisms are utilized. For example, the V_α and J_α gene segment repertoires are very diverse, but the murine T-cell receptor β gene family has a limited number of V_β and D_β gene segments. Nonetheless, the total

germline diversity potential in the V_β genes of normal mice may not be significantly less than that of immunoglobulin V_H and V_κ genes because of the relatively large number of J_β gene segments and the use of D gene segments in three different reading frames. Furthermore, many V_H as well as V_κ gene segments are very closely related and could be nearly equivalent functionally, while the few V_β gene segments are found in a relatively large number of diverse gene segment subfamilies. Some calculations of this diversity-potential for the murine T-cell antigen receptor and for immunoglobulins are presented in Table 3. If we assume that there are 250 V_H gene segments, 10 D_H gene segments, four J_H gene segments, as well as 250 V_κ gene segments, and four functional J_κ gene segments (161), then a total of approximately 10^7 different antibodies could be synthesized. For the T-cell antigen receptor, if we assume that there are 30 V_β gene segments, two D_β gene segments used in all three translational reading frames, 12 J_β gene segments, 100 V_α gene segments, and 50 J_α gene segments, then a total of nearly 10^7 different T-cell antigen receptors are possible as well. Thus, using these reasonable estimates for the number of germline gene segments and assuming that the combinatorial mechanism for generating diversity can operate randomly, the magnitude of germline and combinatorial diversity possible for immunoglobulins and T-cell receptors is similar. Although the number of binding sites that can be generated is similar, clearly the composition of the gene segments involved is different. While the combined immunoglobulin κ and heavy chain families are estimated to have 500 V gene segments and only eight J gene segments, the combined T-cell receptor α and β chains may have nearly as many J gene segments (~65) as V gene segments (~130). Because the 5' ends of the J gene segments are diverse (Figure 6), T-cell receptors will have extra diversity just beyond the V-(D)-J junction, at positions homologous to those that fall within the third immunoglobulin hypervariable region. Thus, diversity is distributed

differently in the variable regions of the calculated T- and B-cell repertoires, although the functional significance of this difference, if any, is unclear.

Perhaps one of the more surprising findings to emerge from the study of T-cell antigen receptors is that somatic hypermutation, a major mechanism for diversity generation in immunoglobulins, does not appear to play a significant role in the T-cell receptor genes. If this is true, then all the mechanisms for diversity generation in these V genes occur early in ontogeny during V gene formation in the thymus. There are three possible explanations for the failure to detect somatic hypermutation in the T-cell receptor genes. The first proposes that somatic hypermutation does occur but is not detected because the sequences analyzed were isolated mostly from cells that have not yet undergone somatic mutation (J. Howard, personal communication). Many of the sequences are derived from β chains synthesized by thymocytes which are relatively immature, and therefore are perhaps not likely to have undergone somatic hypermutation. Moreover, many of the T-cell lines analyzed were stimulated *in vivo* with antigen only once, and were then propagated *in vitro* under antigen-excess conditions, and therefore would not be subject to extensive selection for high-affinity receptor interactions associated with somatic hypermutation. A similar argument can be made for the T-cell hybridomas that do not require antigen to proliferate. Although the maturity of an antigen-stimulated B cell can to some extent be assessed by the immunoglobulin heavy chain isotype expressed (e.g., IgM vs others), there is no similar test for antigen-stimulated T cells. The second explanation proposes that somatic hypermutation in T-cell antigen receptor genes is not permitted because the resulting mutations, especially if they arise after T-cell migration from the thymus, could be deleterious to the organism (203). According to this view, since the T-cell antigen receptor repertoire has been selected for the recognition of antigen in the context of self-MHC-encoded molecules, T cells that undergo

somatic hypermutation in the periphery will generate many autoreactive cells. As T cells on the average are believed to live longer than B cells and T_H cells, unlike B cells, may not dependent on other lymphocytes for activation and proliferation, these autoreactive T-cell somatic mutants present a greater threat than autoreactive B-cells. The third explanation proposes that somatic hypermutation events are rarely detected because of a lack of selection pressure for the expansion of the hypermutation events. This explanation implies that the activation of T cells is not highly dependent upon receptor affinity for antigen or an antigen/MHC complex (188).

MHC-RESTRICTED ANTIGEN RECOGNITION BY T CELLS

One or Two Receptors?

T cells recognize antigen when it is present on the surface of other cells only if the appropriate allele of a polymorphic MHC molecule is also expressed. Thus, T cells have a dual specificity for both antigen and allele-specific determinants of the MHC molecule (1-3), and the recognition of antigen is said to be restricted by the MHC-encoded molecule. This fundamental discovery was made more than a decade ago, and since then a debate has raged over the molecular basis of antigen and MHC recognition by T cells. Two types of models have been proposed to explain this dual specificity. The first hypothesizes that a single T-cell antigen receptor binding site recognizes a combined determinant formed by the restricting MHC molecule and antigen (3, 204-206). This might occur if the antigen and the MHC molecule can interact with sufficient avidity to form a complex antigen. Recent experiments have shown that a peptide fragment of lysozyme capable of stimulating T_H cells can bind to an appropriate responder, but not the nonresponder, I-A molecule with significant affinity (207). The functional relevance of this binding to T-cell activation remains to be determined,

and it is difficult to imagine that a similar, relatively high-affinity binding occurs for each of the antigens that are recognized in the context of a MHC class I or class II molecule. Alternatively, only a low affinity interaction between antigen and the MHC molecule may generally exist which is stabilized by the single binding site of the T-cell receptor (203). The second model proposes that the T-cell receptor is composed of either two separate molecules, or two binding sites in the same molecular complex (208-213). In this case, separate binding sites exist for antigen and the MHC molecule.

Evidence has accumulated against the models in which antigen and MHC are recognized independently. First, independent segregation of antigen and MHC recognition was not found in somatic T-cell hybrids expressing two distinct antigen/MHC specificities (214). Second, when a panel of several hundred T-cell hybridomas was screened for reactivity with a clone-specific antibody that binds a particular α/β heterodimer, a second antibody-reactive clone was found that possessed the same antigen/MHC specificity as well as a fortuitous allo-MHC cross reactivity (215). This observation suggests that the same α/β heterodimer is involved in both antigen and MHC recognition. Finally, for both MHC class I and class II restricted T-cell clones, antigen recognition was shown to be directly influenced by the MHC haplotype expressed by the stimulating or target cells (216-219). For example, several instances where T hybridomas that recognize peptides of cytochrome c along with either of two I-E molecules have been described. The fine specificity of these individual T cells for a set of related cytochrome c peptides depended on which of the two possible I-E restriction elements was used to present antigen. Although for each of these experiments more than one interpretation has been proposed (220), these studies and others taken together imply that antigen and the MHC molecule are not recognized independently (214-219, 221-223), but that dual recognition is mediated through a

receptor composed of α and β subunits. As we shall discuss below, the structure of the T-cell receptor itself argues against a model in which the α and β chains constitute two independent binding sites. This implies that a single receptor binding site can recognize both antigen and MHC.

Despite this evidence, the single-binding site model has not received universal acceptance. The major unresolved difficulties with such models are the means by which MHC molecules can interact with a diversity of antigens to form a single determinant and the means of selection of a self-MHC restricted repertoire of T-cell receptors. Although hypotheses have been offered to account for these problems (204-206, 224, 225), none has won general acceptance. Despite problems with two binding site models, the discovery of the γ chain had lead to renewed speculation on the possibility of a second T-cell antigen receptor (212). The γ chain, although limited in diversity, could associate with the α or β (160, 212) chains or with a diverse, hypothetical δ chain to generate a second T-cell antigen binding site. Alternatively, if the γ chain formed a homodimer, or associated with a monomorphic molecule, it might act as an accessory molecule similar to Lyt-2.

Structural Comparisons Between Immunoglobulin and T-Cell Receptor V Genes

The V and C regions of the α, β, γ and immunoglobulin polypeptides exhibit significant sequence similarity and have descended from common ancestral genes. Since the immunoglobulin molecule has been thoroughly analyzed at the level of primary, secondary and tertiary structure (120, 226), it is pertinent to compare the structures of the T-cell receptor (α/β) and immunoglobulin V regions to gain insights into shared structural features relevant to binding of antigenic determinants.

SECONDARY STRUCTURE As noted above, immunoglobulin V regions fold into two planes of anti-parallel β strands stabilized by the invariant disulfide

bridge. The β strands are relatively conserved in sequence and are connected to one another by polypeptide loops, some of which constitute the hypervariable regions. The secondary structures of V_L and V_H regions are highly conserved. Several algorithms have been developed to assess the tendency of polypeptides of known sequence to have similar secondary structures. First, hydrophobicity analyses measure the hydrophobic characteristics of amino acid side chains (227). From this analysis the exposure of side chains to the interior or exterior of the molecule can be inferred. Second, techniques have been devised for estimating the tendency for short regions of a polypeptide chain to form β-strand structures (228). Analyses of the average hydrophobicity or the β-strand forming potential of large collections of V_α, V_β, V_H and V_κ regions have shown that all four sets of V region are very similar to each other in these properties (87, 88, 94, 95).

AMINO ACID SIMILARITIES Invariant or nearly invariant amino acids are present at eight positions in immunoglobulin variable regions, five in the V segment and three in the J joining region. At these positions, a single amino acid is present in greater than 95% of both light (κ and λ) and heavy chain variable regions from several species (229). Four other positions are semi-invariant, a single amino acid is present in more than 75% of κ, λ and heavy chain variable regions (229). In addition, there are a number of other positions where one set of variable regions, heavy or light, have a particular amino acid, and the other set does not (229). Fourteen of these conserved positions, 11 shared by both V_H and V_L regions, and three in which V_H and V_L regions are invariant or semi-invariant, but differ from one another, are shown in Table 4. These 14 were selected from a larger group of conserved positions by two criteria: (1) the three-dimensional structures of immunoglobulins McPC603 and NEW indicate they might be important for V_H-V_L interactions (230, 231), and (2) the three-dimensional structure of both the NEW light and the NEW heavy chain indicates that these amino

acids form hydrogen bonds that might be important for intrachain folding (231).

Many of the conserved amino acids in immunoglobulin variable regions are also present in the α, β and γ chains. The invariant and semi-invariant amino acids in the β and α V gene segments are indicated above the blocks of sequence in Figures 4 and 5, respectively, and the invariant positions in the J_α, J_β and J_γ gene segments are similarly denoted in Figure 6. Table 4 compares the V_α, V_β and V_γ regions with immunoglobulins at the 14 conserved positions described above. Most of the 14 are present in the three T-cell families of variable regions. For example, all of the eight amino acids that are highly conserved (>95%) in both the V_H and V_L regions are also present in both V_α and V_β regions, and seven are also present in the V_γ region. Seven of these highly conserved positions are indicated by asterisks in Table 4, and the role these play in immunoglobulin intra-V region folding and V_H-V_L interaction is indicated. The presence of these invariant amino acids, and most of the others listed in Table 4, in both V_α, V_β and V_γ regions argues strongly that these variable regions are similar to immunoglobulins in overall structure.

INTERACTIONS BETWEEN V REGIONS The amino acids thought to be important for contact between immunoglobulin V_H and V_L regions include both variable and more conserved amino acids. Those contacts between two conserved amino acids on the light and heavy chain are likely to be most important for understanding the general features of V_H-V_L interactions. If we consider the three-dimensional structures of the mouse myeloma protein McPC603 and the human myeloma protein NEW, five amino acids on both the V_L and V_H regions are involved in such contacts (230, 231). Using the numbering of Kabat *et al.*, these include tyrosine-36L, glutamine-38L, proline-44L, tyrosine-87L, and phenylalanine-98L in both κ and λ light chains, and valine-37H, glutamine-39H, leucine-45H, tyrosine-91H, and tryptophan-103H in the heavy chain. Several of

these amino acids are also highly conserved in α and β variable regions. The glutamine is found at the homologous position in both V_α and V_β regions. The V_β regions have leucine and the V_α regions have either proline or leucine at a position homologous to leucine 45H or proline 44L. On the other hand, the tyrosine (87L or 91H) is not commonly found in α and β variable regions which often have a conservative substitution of phenylalanine or leucine at this position (Table 4). At the two other conserved positions important for interchain contacts, both the V_α and V_β regions tend to have a tyrosine and a phenylalanine, and therefore resemble V_L regions somewhat more than they resemble V_H regions, which nearly always have valine and tryptophan at these positions (95) (Table 4). Furthermore, the distance between the cysteines that form the intrachain disulfide bond of the V_α and the V_β regions is 63-69 amino acids (95). Again, this property is more similar to the V_L regions where this distance is 64-69 amino acids, than it is to V_H regions, where it is 69-75 amino acids (229). The α/β heterodimer may therefore resemble a light chain dimer more than it does a heavy-light chain pair. This is intriguing, because the structure of one light chain dimer that has been studied by X-ray analyses is unusual in that it has a very deep antigen-binding pocket (232). However, the tendency for V_α and V_β regions to be more similar to V_L regions is not uniform. In the length of their J gene segments and the presence of conserved amino acids homologous to alanine-92H and valine 111-H, the V_α and the V_β regions are more similar to V_H than V_L regions. It is therefore likely that the slightly greater resemblance of the V_α and the V_β regions to the V_L regions is due to convergent rather than divergent evolution of these V regions.

Antigen Binding by the α/β Heterodimer

HYPERVARIABLE REGIONS Calculations of variability, defined as the number of different amino acids occurring at a given position divided by the frequency of the most common amino acid at that position, permit the assessment

of diversity at each position in a set of variable regions (233). An analysis of V_L and V_H regions with this approach reveals three distinct hypervariable segments—one around positions 24-35, a second near positions 50-65, and a third at the V-J or V-D-J junctions, or positions 89-102 (233). The positions of the hypervariable regions correspond to those portions of the molecule that X-ray crystallographic studies have shown to be in contact with antigens (120, 226, 233). Examination of V_L and V_H region hypervariable plots also reveals some additional regions of variability around positions 10-20 and 75-80, particularly in light chains, although these are not as variable as the three "classical" hypervariable regions (229). The identification of hypervariable regions of T-cell receptor α and β chains are therefore expected to indicate those portions of the molecule involved in antigen/MHC contact. Such analyses have been carried out, and several general conclusions can be drawn. First, in the β chain, there is greater variability in positions homologous to those previously defined as hypervariable in immunoglobulins (87-89). Second, there is an overall higher level of background variability in both α and β chains, than in immunoglobulin light and heavy chains (87-89, 94, 95). Interestingly, hypervariability plots of collections of V_α regions have shown the variability to be broadly distributed throughout the variable region with the exception of the third hypervariable region, which is more variable, and a few scattered conserved positions (B. Arden, unpublished observation). This broader distribution of variability in both chains is consistent with other comparisons indicating that there are fewer highly conserved amino acids in T-cell receptor V regions and that the average V_α and V_β regions are more different from one another than are immunoglobulin V regions (M. Kronenberg, unpublished observation and refs. 87-89). However, the immunoglobulin sample used for comparison contains many myeloma proteins and hapten-binding proteins, and there is some evidence these may not represent the full extent of the

diversity of the germline immunoglobulin repertoire (234-236). Third, it has been proposed that the V_β gene segments have several extra distinct hypervariable regions (87), although this was not observed in a second analysis (88). The discrepancy is best explained by differences in sequence alignments and the sensitivity of the analysis to slight sample differences when relatively small samples are employed.

Because of the apparent greater variability of V_α and V_β regions, and the possible existence of extra β chain hypervariable regions, it has been hypothesized that for the T-cell receptor a greater surface interacts with antigen than is the case for immunoglobulin (87). Although this is plausible, considering that the T-cell receptor is required to bind a complex determinant containing both antigen and MHC, there is too little structural information on the interaction of antibodies with protein antigens, as opposed to haptens, to definitively state that the two sets of receptors must be different from one another. The available crystallographic structure of an antibody binding lysozyme suggests that a large surface might be involved in binding (237). This type of antibody-antigen interaction may account for the extra regions of immunoglobulin variability mentioned above and may well be comparable to the T-cell receptor interaction with antigen in the context of MHC.

THE α/β HETERODIMER: ONE OR TWO SITES? In summary, the V_α, V_β, V_H and V_L regions all appear to exhibit a conserved set of amino acids that are important in immunoglobulins for V_H-V_L interactions, a striking similarity in their patterns of secondary structure and similarity in their variability patterns. The data suggest that the antibody fold is conserved in the V_α and the V_β regions and that the V_α and V_β regions interact like in antibodies to generate a single binding site. In this regard, immunoglobulins apparently can be MHC restricted; several antibodies that recognize influenza polypeptides in the context of a specific MHC

allele have been described (238). Accordingly, the antibody binding site has at least the potential for recognizing antigen in the context of an MHC molecule, and it is therefore possible that a single T-cell receptor could do the same. Alternatively, the α/β heterodimer may be responsible only for antigen or MHC recognition, in which case a second receptor needs to be defined.

Our view that the V_α and V_β regions fold similar to immunoglobulins to form a single binding site is not consistent with models that require the V_α and the V_β regions to form separate antigen and MHC binding sites, either as single chains (220), or in association with other molecules such as individual subunits of the Ia molecule (239). However, this single-site view of the α/β heterodimer can be easily reconciled with the lack of correlation between the expression of particular V_β gene segments and the ability to recognize either particular antigens or MHC molecules (Table 5). In one of the earliest studied cases, it was found that a T-cell hybridoma specific for cytochrome c and I-E^k and a T-cell line specific for lysozyme and I-A^b both express the $V_{\beta 3}$ gene segment (148) (Figure 12); however, other cytochrome c/I-E^k-specific hybridomas do not use this V_β gene segment (J. Goverman and A. Winoto, unpublished). Since then a number of similar cases have been described (Table 5). Because in immunoglobulins specificity can be associated with use of a particular light chain (240), a particular heavy chain (241), or both, we expect that the same will hold true for the T-cell receptor and that valid generalizations about α or β usage and antigen or MHC specificity will not be possible.

MHC Class I- and Class II-Specific T Cells

There is a correlation between T-cell recognition of MHC class I molecules, expression of γ gene RNA and cytotoxic function on the one hand (149, 242-244), and a correlation between T-cell recognition of antigens in the context of MHC class II molecules and helper function and/or IL-2 secretion on the other (242,

245-248). The reason for these correlations is unknown, although the structure of the T-cell antigen receptor in the two categories of cells is probably not responsible. Both T_H and T_C express C_α and C_β RNA, indicating that unlike immunoglobulins there are not function-associated T-cell receptor constant region isotypes. Furthermore, the T-cell receptors for both murine T_H and T_C cells appear to be encoded by the same pool of V_β, D_β and J_β gene segments. For example, analyses of a T-cell hybridoma specific for TNP and I-A^d and a cloned T-cell line specific for an H-2^d class I molecule have shown that both utilize the $V_{\beta 2}$ gene segment in their antigen receptor (88) (Table 5). In addition, two monoclonal antibodies have been generated that each recognize a single V_β gene segment subfamily. KJ16-133 (52) is specific for the murine $V_{\beta 8}$ subfamily, which contains three members (87, 88), while Ti_{3A} (194) is specific for the human $V_{\beta M3}$ subfamily, which contains five members, three of which are functional (93). Both monoclonal antibodies bind to both L3T4 or T4 positive populations enriched for helper T cells and Lyt-2 or T8 positive T-cell populations enriched for cytotoxic T cells. Therefore cells of both phenotypes utilize the same V_β gene segments (179, 194). There is currently no data indicating whether the same might also be true for V_α gene segment usage.

Development of the T-Cell Specificity Repertoire

The ontological origin of a self-MHC restricted, antigen-specific repertoire of T-cell receptors has long been a controversial issue. It is now clear from experiments that determined the chromosomal locations of the α and β genes (Table 1) that these genes are not linked to those encoding the MHC molecules. It is therefore intriguing that MHC-congenic strains of mice with the same germline receptor genes express antigen-specific repertoires that in each strain are strongly biased for the recognition of different self-MHC molecules.

A number of studies, including those utilizing chimeric mice, have explored this question (249-251). Chimeric mice are constructed for this purpose by transferring precursor stem cells from bone marrow of a mouse with one MHC genotype into another lethally irradiated mouse having a different MHC genotype. Stem cells in the bone marrow repopulate the hematopoietic system and therefore in the chimera the hematopoietic cells have a different genotype than the irradiated recipient. Alternatively, chimeras have been constructed by transplanting thymus tissue into athymic mice that have a different MHC genotype. Many, but not all, of the chimera studies are consistent with a "thymic education" model for the ontogeny of a self-MHC restricted repertoire. According to this model, T cells with particular antigen-receptors are selected for in the thymus in the absence of antigen. This selection is believed to depend upon MHC molecules expressed by certain thymic cells, such that T cells capable of recognizing antigens in the context of self-MHC molecules emigrate and populate the peripheral lymphoid organs. T cells that only recognize antigen plus allo-MHC, or that recognize antigen alone, do not emigrate, and are presumed to die in the thymus. This selection process therefore provides a rationale for the extensive cell death that takes place in thymus. According to several models of this type, selection is associated with a somatic hypermutation process (209, 210). The thymic education model contrasts with alternative antigen-priming models which assert that selection of T-cell specificities restricted by a given MHC haplotype occurs only in the periphery during T-cell activation in the presence of antigen (250). According to these models, mature cells emerging from the thymus are not particularly biased in favor of recognizing antigen plus self-MHC as opposed to antigen plus allo-MHC. However in the peripheral lymphoid organs, the combination of antigen plus the available MHC molecules, will lead to the clonal expansion of those T cells restricted to self-MHC molecules.

Regulatory influences by suppressor and other cells may also prevent expansion of allo-MHC restricted clones.

Although study of the molecular biology of the α and β genes has not resolved this controversy, the available evidence permits the following conclusions. First, the thymus is the site where the α/β heterodimer is first expressed. Therefore selection of receptor specificities or induction of tolerance before the entry of T cells into the thymus is unlikely. These data do not argue directly for thymic selection as opposed to selection in peripheral lymphoid organs. Second, as mentioned above, up to 70% of the cell-death in the thymus may be caused by nonproductive α and β gene rearrangements. Some additional cell-death may also be due to the removal of self-reactive cells. However these two explanations do not account quantitatively for the significant amount of cell death, estimated to be up to 95% of the total thymocytes. If the estimate is accurate there must be additional causes for this phenomena (252-254). Finally, the absence of conclusive evidence for somatic hypermutation in both the α and β genes of the receptors expressed by peripheral T cells rules out models that require this as an obligatory step in the selection of a self-MHC restricted T-cell repertoire. Therefore the postulated selection of T-cell receptor specificities must act only upon the expressed receptor formed by germline genes and the diversity-generating rearrangement processes described above.

Implications for Immune Response Gene Defects

Immune response (Ir) gene defects are said to occur when heritable differences in responsiveness to the same antigen are demonstrated among individuals. In such cases, one group or individual are responders, while the other group manifesting the so-called "defect" are nonresponders (250, 255, 256). In most cases, it has been shown that responsiveness in inbred mice maps to MHC class I molecules for cytotoxic cells (257), and to class II molecules for helper T-

cell responses (258). In these cases, nonresponsiveness is highly specific for the combination of particular antigens and particular alleles of MHC-encoded molecules. Several models have been proposed to explain these Ir gene-encoded immune response gene defects (250, 255). One possibility is that germline genes that could encode a receptor with the specificity in question do not exist. A second explanation for this selective constraint proposes that selection of a self-MHC restricted repertoire in the absence of antigen (see above), leads incidentally to deletion of clones reactive with certain antigen/MHC combinations. The exact details of the relationship between repertoire selection and clonal deletion depend on whether one believes in one or two site models of the receptor (209, 210, 256). Both of these explanations imply that T cell responses that show Ir gene control probably require the expression of relatively few genes, so that occasionally certain strains have germline deletions that include these genes such that gaps in the repertoire can be detected, or alternatively that selection can reveal these gaps. A second category of models propose that selective constraints lead to an absence of clones that recognize the combination of antigen plus MHC molecule. The requirement for self-tolerance is an example of such a selective constraint. Deletion of self-reactive clones could eliminate a significant portion of the potential repertoire, including all those T cells that react with certain foreign antigen/self MHC combinations if the receptor expressed by these cells fortuitously cross-reacts with a self-determinant (259).

An alternative hypothesis for Ir gene defects does not invoke an inability to express a particular T-cell specificity, but proposes instead that certain antigen/MHC combinations may not effectively activate T cells in the periphery, although T cells that could react with these determinants are present. For example, the determinant selection theory states that a single T-cell receptor recognizes an antigen-MHC complex, but that in some cases a particular MHC molecule and

antigen may fail to interact with the avidity required to form a complex and stimulate T lymphocytes (225). Another hypothesis invokes regulatory interactions between potentially responsive clones and suppressor cells that prevent responsiveness.

Two recent observations favor those explanations that do not require an unavailability or deletion of clones expressing certain specificities. First, recent studies characterizing the specificity profiles of T-cell clones reactive to several small peptides have revealed considerable diversity among clones. For example, T-cell clones derived from B10.A mice immunized with a 23 amino acid peptide of lysozyme (amino acids 74-96) showed a strict correlation between the minimal peptide determinants and the Ia molecule required for recognition; namely, clones specific for amino acids 74-86 were always I-A^k restricted, and clones reacting with amino acids 85-96 were I-E^k restricted (260 and N. Shastri, G. Gammon, A. Miller, and E. Sercarz, unpublished). This correlation between the antigenic determinant and the MHC molecule also illustrates the phenomenon of Ir gene defects in that B10.A mice are nonresponders to both peptide 74-86/I-E^k as well as to peptide 85-96/I-A^k. Further analysis of the specificity profiles of these peptide-specific clones has shown that among each of the two sets restricted to different class II molecules, individual clones can be distinguished by their reactivities to a set of variant synthetic peptides. It therefore appears that T cells can recognize these small peptides along with the restricting MHC molecule, in several different ways.

The diversity of T-cell specificity phenotypes observed for even minimal peptide determinants argues against absence of appropriate clones, or clonal-deletion mechanisms as valid explanations for Ir gene defects. Considering for example the self-tolerance model, given the diversity of the observed responses, it is difficult to see how fortuitous cross-reactions with a self antigen could cause

deletion of all the potential T cells reactive with the peptide. This could occur if the observed distinct specificity patterns of the clones were due to the fact that these clones had descended from a single parental clone through a somatic hypermutation mechanism. However, the sequencing of the V_β genes of the receptors employed by these clones has failed to reveal any somatic hypermutation, and different V_β sequences were found for each of the seven I-A^k or I-E^k restricted clones analyzed (J. Kobori & N. Shastri, unpublished). Similar diversity has been found for the β genes expressed in a series of cytochrome c-specific T-cell hybrids (A. Winoto, N. Lan & D. Hansburg, unpublished data). This demonstrates that Ir gene regulated T-cell responses directed towards even minimal peptide determinants are not constrained in their usage of particular V_β genes, and argues against the models described above which predict that a relatively limited set of genes would be employed in these responses.

Another observation that argues against absence of clones having certain T-cell receptors as the cause of Ir gene defects is the finding that in one instance immune responsiveness to a different lysozyme peptide (amino acids 46-61) could be clearly correlated with binding of this peptide to an appropriate, but not to an inappropriate or nonresponder MHC class II molecule. This result is most consistent with the determinant selection hypothesis, although as mentioned previously, the generality of this finding has not yet been established.

In summary, these findings argue against a limitation in the existence or expression of appropriate T-cell receptor genes as a complete explanation for nonresponsiveness. To some extent, the studies cited support the alternative explanations including determinant selection and regulation by suppressor cells that assign the Ir defect to the activation of T cells in the periphery.

SUMMARY

The genes encoding the α and β chain of the T-cell receptor and the γ gene have been cloned, and their structure, organization, ontogeny of expression, pattern of rearrangement and diversification are now generally understood. In most cases, the immunoglobulin paradigm applied very well to the corresponding phemenona in T cells, although as described above, some interesting and potentially important differences exist. Nevertheless, there are still many unanswered questions regarding the ontogeny and mechanism of MHC-restricted antigen recognition, and it is not clear how far the immunoglobulin model can take us in understanding these phenomena. Although the α/β heterodimer looks like an antibody and the binding sites of the two molecules may be similar, the rules governing B- and T-cell activation are clearly different and the ligand(s) bound by the receptor are still poorly characterized. In the future T-cell receptor genes as well as those encoding the T-cell accessory molecules will be altered *in vitro* and transferred into mammalian cells in culture and into whole organisms in an attempt to understand T-cell antigen recognition. These tools will allow us to manipulate the mammalian immune response in a variety of different ways that will have a profound impact on both our understanding of immunology and on medicine in the future.

Acknowledgements. The literature search for this article was completed in July, 1985. We thank P. Pjura for computer assistance in visualizing three-dimensional immunoglobulin structures, Drs. A. Augustin, M. Cohn, M. Davis, E. Haber, J. Kappler, R. Langman, D. Loh, T. Mak, P. Marrack, E. Palmer, T. Rabbitts, G.-K. Sim, S. Tonegawa & E. Unanue for sending us their papers prior to publication, Drs. Richard Barth, Ulf Landegren and Astar Winoto for critical reading of the

manuscript, our colleagues at Caltech for their many helpful suggestions, and S. Mangrum, R. Thorf, K. Patterson, C. Katz, and B. Jones for helping with all the stupid drafts and figures.

Literature Cited

1. Kindred, B., Shreffler, D. C. 1972. H-2 dependence of cooperation between T and B cells in vivo. *J. Immunol.* 109:940-943.

2. Katz, D. H., Hamaoka, T., Benacerraf, B. 1973. Cell interactions between histoincompatible T and B lymphocytes. II. Failure of physiologic cooperative interactions between T and B lymphocytes from allogeneic donor strains in humoral response to hapten-protein conjugates. *J. Exp. Med.* 137:1405-1418.

2a. Rosenthal, A. S., Shevach, E. M. 1973. Function of macrophages in antigen recognition by guinea pig T lymphocytes. *J. Exp. Med.* 138:1194-1212.

3. Zinkernagel, R. M., Doherty, P. C. 1974. Restriction of in vitro T-cell mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or semi-allogeneic system. *Nature* 248:701-702.

4. Cerottini, J. C., Nordin, A. A., Brunner, K. T. 1970. Specific *in vitro* cytotoxicity of thymus-derived lymphocytes sensitized to alloantigens. *Nature* 228:1308-1309.

5. Mitchell, G. F., Miller, J. F. A. P. 1968. Cell-to-cell interaction in the immune response. II. The source of hemolysin-forming cells in irradiated mice given bone marrow and thymus or thoracic duct lymphocytes. *J. Exp. Med.* 128:821-837.

6. Cantor, H., Boyse, E. A. 1975. Functional subclasses of T lymphocytes bearing differing Ly antigens. II. Cooperation between subclasses of Ly^+ cells in the generation of killer activity. *J. Exp. Med.* 141:1390-1399.

7. Gershon, R. K. 1974. T-cell suppression. *Contemporary Topics in Immunobiology* 3:1-40.

8. Cantor, H., Weissman, I. 1976. Development and function of subpopulations of thymocytes and T lymphocytes. *Progress in Allergy* 20:1-64.

9. Whitlock, C., Denis, K., Robertson, D., Witte, O. 1985. In vitro analysis of murine B-cell development. *Ann. Rev. Immunol.* 3:213-235.

10. Weissman, I. L. 1967. Thymus cell migration. *J. Exp. Med.* 126:291-304.

11. Scollay, R., Bartlett, P., Shortman, K. 1984. T-cell development in the adult murine thymus: changes in the expression of the surface antigens Ly2, L3T4 and B2A2 during development from early precursor cells to emigrants. *Immunol. Rev.* 82:79-103.

12. Rothenberg, E., Lugo, J. Differentiation and cell division in the mammalian thymus. *Devel. Biol.*, in press.

13. Borst, J., Alexander, S., Elder, J., Terhorst, C. 1983. The T3 complex on human T lymphocytes involves four structurally distinct glycoproteins. *J. Biol. Chem.* 258:5135-5143.

14. Meuer, S. C., Cooper, D. A., Hodgdon, J. C., Hussey, R. E., Fitzgerald, K. A., Schlossman, S., Reinherz, E. L. 1983. Identification of the receptor for antigen and major histocompatibility complex on human inducer T lymphocytes. *Science* 222:1239-1242.

15. Engleman, E. G., Benike, C., Glickman, E., Evans, R. L. 1981. Antibodies to membrane strutures that distinguish suppressor/cytotoxic and helper T lymphocyte subpopulations block the mixed leukocyte reaction in man. *J. Exp. Med.* 154:193-198.

16. Landegren, U., Romstedt, U., Axberg, I., Ullberg, M., Jondal, M., Wigzell, H. 1982. Selective inhibition of human T-cell cytotoxicity at levels of target recognition or initiation of lysis by monoclonal OKT3 and Leu-2a antibodies. *J. Exp. Med.* 155:1579-1584.

17. Marrack, P., Endres, R., Shimonkevitz, R., Zlotnik, A., Dialynas, D., Fitch, F., Kappler, J. 1983. The major histocompatibility complex-restricted antigen receptor on T cells. II. Role of the L3T4 product. *J. Exp. Med.* 158:1077-1091.

18. Swain, S. L. 1981. Significance of Lyt phenotypes: Lyt2 antibodies block activities of T cells that recognize class I major histocompatibility complex antigens regardless of their function. *Proc. Natl. Acad. Sci. USA* 78:7101-7105.

19. Biddison, W. E., Rao, P. E., Talle, M. A., Goldstein, G., Shaw, S. 1984. Possible involvement of the T4 molecule in T-cell recognition of class II HLA antigens: evidence from studies of CTL-target cell binding. *J. Exp. Med.* 159:793-797.

20. Allison, J. P., McIntyre, B. W., Bloch, D. 1982. Tumor-specific antigen of murine T lymphoma defined with monoclonal antibody. *J. Immunol.* 129:2293-2300.

21. Haskins, K., Kubo, R., White, J., Pigeon, M., Kappler, J., Marrack, P. 1983. The major histocompatibility complex-restricted antigen receptor on T cells. I. Isolation with a monoclonal antibody. *J. Exp. Med.* 157:1149-1169.

22. Lancki, D. W., Lorber, M. I., Loken, M. R., Fitch, F. W. 1983. A clone-specific monoclonal antibody that inhibits cytolysis of a cytolytic T-cell clone. *J. Exp. Med.* 157:921-935.

23. McIntyre, B. W., Allison, J. P. 1983. The mouse T-cell receptor: structural heterogeneity of molecules of normal T cells defined by xenoantiserum. *Cell* 34:739-746.

24. Meuer, S. C., Fitzgerald, K. A., Hussey, R. E., Hodgdon, J. C., Schlossman, S. F., Reinherz, E. L. 1983. Clonotypic structures involved in antigen-specific human T-cell function. Relationship to the T3 molecular complex. *J. Exp. Med.* 157:705-719.

25. Samelson, L. E., Germain, R. N., Schwartz, R. H. 1983. Monoclonal antibodies against the antigen receptor on a cloned T-cell hybrid. *Proc. Natl. Acad. Sci. USA* 80:6972-6976.

26. Haars, R., Rohowsky-Kochan, C., Reed, E., King, D. W., Suciu-Foca, N. 1984. Modulations of T-cell antigen receptor on lymphocyte membrane. *Immunogenetics* 20:397-405.

27. Kappler, J., Kubo, R., Haskins, K., Hannum, C., Marrack, P., Pigeon, M., McIntyre, B., Allison, J., Trowbridge, I. 1983. The major histocompatibility complex-restricted antigen receptor on T cells in mouse and man: identification of constant and variable peptides. *Cell* 35:295-302.

28. Meuer, S. C., Acuto, O., Hercend, T., Schlossman, S. F., Reinherz, E. L. 1984. The human T-cell receptor. *Ann. Rev. Immunol.* 2:23-50.

29. Yanagi, Y., Yoshikai, Y., Leggett, K., Clark, S. P., Aleksander, I, Mak, T. W. 1984. A human T cell-specific cDNA clone encodes a protein having extensive homology to immunoglobulin chains. *Nature* 308:145-149.

30. Hedrick, S. M., Cohen, D. I., Nielsen, E. A., Davis, M. M. 1984. Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* 308:149-153.

31. Saito, H., Kranz, D. M., Takagaki, Y., Hayday, A. C., Eisen, H. N., Tonegawa, S. 1984. Complete primary structure of a heterodimeric T-cell receptor deduced from cDNA sequences. *Nature* 309:757-762.

32. Chien, Y., Becker, D., Lindsten, T., Okamura, M., Cohen, D., Davis, M. 1984. A third type of murine T-cell receptor gene. *Nature* 312:31-35.

33. Saito, H., Kranz, D., Takagaki, Y., Hayday, A., Eisen, H., Tonegawa, S. 1984. A third rearranged and expressed gene in a clone of cytotoxic T lymphocytes. *Nature* 312:36-40.

34. Hedrick, S. M., Nielsen, E. A., Kavaler, J., Cohen, D. I., Davis, M. M. 1984. Sequence relationships between putative T-cell receptor polypeptides and immunoglobulins. *Nature* 308:153-158.

35. Fabbi, M., Acuto, O., Smart, J. E., Reinherz, E. L. 1984. Homology of Ti alpha-subunit of a T-cell antigen-MHC receptor with immunoglobulin. *Nature* 312:269-271.

36. Hannum, C. H., Kappler, J. W., Trowbridge, I. S., Marrack, P., Freed, J. H. 1984. Immunoglobulin-like nature of the α-chain of a human T-cell antigen/MHC receptor. *Nature* 312:65-67.

37. Sim, G. K., Yagüe, J., Nelson, J., Marrack, P., Palmer, E., Augustin, A., Kappler, J. 1984. Primary structure of human T-cell receptor α-chain. *Nature* 312:771-775.

38. Nathans, J., Hogness, D. S. 1983. Isolation, sequence analysis, and intron-exon arrangement of the gene encoding bovine rhodopsin. *Cell* 34:807-814.

39. Dembic, Z., Bannwarth, W., Taylor, B. A., Steinmetz, M. 1985. The gene encoding the T-cell receptor α-chain maps close to the Np-2 locus on mouse chromosome 14. *Nature* 314:271-273.

40. Kranz, D. M., Saito, H., Disteche, C. M., Swisshelm, K., Pravtcheva, D., Ruddle, F. H., Eisen, H. N., Tonegawa, S. 1985. Chromosomal locations of the murine T-cell receptor alpha-chain gene and the T-cell gamma gene. *Science* 227:941-945.

41. Rabbitts, T. H., Lefranc, M. P., Stinson, M. A., Sims, J. E., Schroeder, J., Steinmetz, M., Spurr, N. L., Solomon, E., Goodfellow, P. N. 1985. The chromosomal location of T-cell receptor genes and a T-cell rearranging gene: possible correlation with specific translocations in human T-cell leukaemia. *EMBO J.* 4:1461-1465.

42. Roehm, N. W., Carbone, C., Kushnir, E., Taylor, B. A., Riblet, R. J., Marrack, P., Kappler, J. W. 1985. The major histocompatibility complex-restricted antigen receptor on T cells: the genetics of expression of an allotype. *J. Immunol.* 135:2176-2182.

43. Epstein, R., Roehm, N., Marrack, P., Kappler, J., Davis, M., Hedrick, S., Cohn, M. 1985. Genetic markers of the antigen-specific T-cell receptor locus. *J. Exp. Med.* 161:1219-1224.

44. Barker, P. E., Ruddle, F. H., Royer, H. D., Acuto, O., Reinherz, E. L. 1984. Chromosomal location of human T-cell receptor gene Ti β. *Science* 226:348-349.

45. Caccia, N., Kronenberg, M., Saxe, D., Haars, R., Bruns, G., Goverman, J., Malissen, M., Willard, H., Yoshikai, Y., Simon, M., Hood, L., Mak, T. 1984. The T-cell receptor β chain genes are located on chromosome 6 in mice and chromosome 7 in humans. *Cell* 37:1091-1099.

46. Morton, C. C., Duby, A. D., Eddy, R. L., Shows, T. B., Seidman, J. G. 1985. Genes for β chain of human T-cell antigen receptor map to regions of chromosomal rearrangement in T cells. *Science* 228:582-585.

47. Lee, N. E., D'Eustachio, P., Pravtcheva, D., Ruddle, F. H., Hedrick, S. M., Davis, M. M. 1984. Murine T-cell receptor β chain is encoded on chromosome 6. *J. Exp. Med.* 160:905-913.

48. Hengartner, H., Meo, T., Muller, E. 1978. Assignment of genes for immunoglobulin *k* and heavy chains to chromosomes 6 and 12 in the mouse. *Proc. Natl. Acad. Sci. USA* 75:4494-4498.

49. Swan, D., Eustachio, P. D., Leinwand, L., Seidman, J., Keithley, D., Ruddle, F. H. 1979. Chromosomal assignment of the mouse κ light chain genes. *Proc. Natl. Acad. Sci. USA* 76:2735-2739.

50. Itakura, K., Hutton, J. J., Boyse, E. A., Old, L. J. 1972. Genetic linkage relationships of loci specifying differentiation alloantigens in the mouse. *Transplantation* 13:239-243.

51. Gottleib, P. D., Durda, P. J. 1976. The I_{β}-peptide marker and the Ly-3 surface alloantigen: structural studies of V_{κ}-region polymorphism and a T-cell marker determined by linked genes. *Cold Spring Harbor Symp. Quant. Biol.* 41:805-815.

52. Haskins, K., Hannum, C., White, J., Roehm, N., Kubo, R., Kappler, J., Marrack, P. 1984. The major histocompatibility complex-restricted antigen receptor on T cells. VI. An antibody to a receptor allotype. *J. Exp. Med.* 160:452-471.

53. Sim, G.-K., Augustin, A. 1985. V_{β} gene polymorphism and a major polyclonal T-cell receptor idiotype. *Cell* 42:89-92.

54. Caccia, N., Bruns, G. A. P., Kirsch, I. R., Hollis, G. F., Bertness, V., Mak, T. W. 1985. T-cell receptor α chain genes are located on chromosome 14 at 14q11-14q12 in humans. *J. Exp. Med.* 161:1255-1260.

55. Collins, M. K. L., Goodfellow, P. N., Spurr, N. K., Solomon, E., Tanigawa, G., Tonegawa, S., Owen, M. J. 1985. The human T-cell receptor α-chain gene maps to chromosome 14. *Nature* 314:273-274.

56. Croce, C. M., Isobe, M., Palumbo, A., Puck, J., Ming, J., Tweardy, D., Erikson, J., Davis, M., Rovera, G. 1985. Gene for α-chain of human T-cell receptor: location on chromosome 14 region involved in T-cell neoplasms. *Science* 227:1044-1047.

57. Jones, C., Morse, H. G., Kao, F.-T., Carbone, A., Palmer, E. 1985. Human T-cell receptor α-chain genes: location on chromosome 14. *Science* 228:83-85.

58. Kirsch, I. R., Morton, C. C., Nakahara, K., Leder, P. 1982. Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes. *Science* 216:301-303.

59. Hecht, F., Morgan, R., Hecht, R., Smith, S. D. 1984. Common region on chromosome 14 in T-cell leukemia and lymphoma. *Science* 226:1445-1447.

60. Williams, D. L., Look, A. T., Melvin, S. L., Roberson, P. K., Dahl, G., Flake, T., Stass, S. 1984. New chromosomal translocations correlate with specific immunophenotypes of childhood acute lymphoblastic leukemia. *Cell* 36:101-109.

61. Zech, L., Gahrton, L., Hammarstrom, L., Juliusson, G., Mellstedt, H., Robert, K. H., Smith, C. I. E. 1984. Inversion of chromosome 14 marks human T-cell chronic lymphocytic leukemia. *Nature* 308:858-860.

62. Klein, G. 1983. Specific chromosomal translocations and the genesis of B-cell-derived tumors in mice and men. *Cell* 32:311-315.

63. Leder, P., Battey, J., Lenoir, G., Moulding, C., Murphy, W., Potter, H., Stewart, T., Taub, R. 1983. Translocations among antibody genes in human cancer. *Science* 222:765-771.

64. Collins, M. K. L., Goodfellow, P. N., Dunne, M. J., Spurr, N. K., Solomon, E., Owen, M. J. 1984. A human T-cell antigen receptor β chain gene maps to chromosome 7. *EMBO J.*, 3:2347-2349.

65. Isobe, M., Erikson, J., Emanuel, B. S., Nowell, P. C., Croce, C. M. 1985. Location of gene for β subunit of human T-cell receptor at band 7q35, a region prone to rearrangements in T cells. *Science* 228:580-582.

66. Lebeau, M. M., Diaz, M. O., Rowley, J. D., Mak, T. W. 1985. Chromosomal localization of the human T-cell receptor β-chain genes. *Cell* 41:335.

67. McBride, O. W., Hieter, P. A., Hollis, G. F., Swan, D., Otey, M. C., Leder, P. 1982. Chromosomal location of human kappa and lambda immunoglobulin light chain constant region genes. *J. Exp. Med.* 155:1480-1490.

68. Aurias, A., Dutrillaux, B., Buriot, D., Lejeune, J. 1980. High frequencies of inversions and translocations of chromosomes 7 and 14 in ataxia telangiectasia. *Mutation Res.* 69:369-374.

69. Welch, J. P., Lee, C. L. Y. 1975. Non-random occurrence of 7-14 translocations in human lymphocyte cultures. *Nature* 255:241-242.

70. Ayme, S., Mattei, J. F., Mattei, M. G., Aurran, Y., Giraud, F. 1976. Nonrandom distribution of chromosome breaks in cultured lymphocytes of normal subjects. *Hum. Genet.* 31:161-175.

71. Chien, Y.-H., Gascoigne, N. R. J., Kavaler, J., Lee, N. E., Davis, M. M. 1984. Somatic recombination in a murine T-cell receptor gene. *Nature* 309:322-326.

72. Siu, G., Clark, S., Yoshikai, Y., Malissen, M., Yanagi, Y., Strauss, E., Mak, T., Hood, L. 1984. The human T-cell antigen receptor is encoded by variable, diversity and joining gene segments that rearrange to generate a complete V gene. *Cell* 37:393-401.

73. Hayday, A. C., Saito, H., Gillies, S. D., Kranz, D. M., Tanigawa, G., Eisen, H. N., Tonegawa, S. 1985. Structure, organization and somatic rearrangement of T-cell gamma genes. *Cell* 40:259-269.

74. Hayday, A., Diamond, D., Tanigawa, G., Heilig, J., Folsom, V., Saito, H., Tonegawa, S. 1985. Unusual features of the organization and diversity of T-cell receptor α chain genes. *Nature* 316:828-832.

75. Winoto, A., Mjolsness, S., Hood, L. 1985. Genomic organization of the genes encoding mouse T-cell receptor α chain. *Nature* 316:832-836.

76. Yoshikai, Y., Clark, S. P., Taylor, S., Sohn, U., Wilson, B., Minden, M., Mak, T. W. 1985. Organization and sequences of the variable, joining and constant region genes of the human T-cell receptor α chain. *Nature* 316:837-840.

77. Gascoigne, N. R. J., Chien, Y.-H., Becker, D. M., Kavaler, J., Davis, M. M. 1984. Genomic organization and sequence of T-cell receptor β-chain constant- and joining-region genes. *Nature* 310:387-391.

78. Malissen, M., Minard, K., Mjolsness, S., Kronenberg, M., Goverman, J., Hunkapiller, T., Prystowsky, M., Yoshikai, Y., Fitch, F., Mak, T., Hood, L. 1984. Mouse T-cell antigen receptor: structure and organization of constant and joining gene segments encoding the β polypeptide. *Cell* 37:1101-1110.

79. Yanagi, Y., Chan, A., Chin, B., Minden, M., Mak, T. W. 1985. Analysis of cDNA clones specific for human T cells and the α and β chains of the T-cell receptor heterodimer from a human T-cell line. *Proc. Natl. Acad. Sci.* 82:3430-3434.

80. Clark, S. P., Yoshikai, Y., Taylor, S., Siu, G., Hood, L., Mak, T. W. 1984. Identification of a diversity segment of the human T-cell receptor β chain, and comparison to the analogous murine element. *Nature* 311:387-389.

81. Kavaler, J., Davis, M. M., Chien, Y.-H. 1984. Localization of a T-cell receptor diversity region element. *Nature* 310:421-423.

82. Siu, G., Kronenberg, M., Strauss, E., Haars, R., Mak, T., Hood, L. 1984. The structure, rearrangement and expression of D_β gene segments of the murine T-cell antigen receptor. *Nature* 311:344-350.

83. Max, E. E., Seidman, J. G., Leder, P. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin κ constant region. *Proc. Natl. Acad. Sci.* 76:3450-3454.

84. Sakano, H., Hüppi, K., Heinrich, G., Tonegawa, S. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* 280:288-294.

85. Early, P., Huang, H., Davis, M., Calame, K., Hood, L. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. *Cell* 19:981-992.

86. Malissen, M., McCoy, C., Blanc, D., Trucy, J., Devaux, C., Schmitt-Verhulst, A.-M., Fitch, F., Hood, L., Malissen, B. 1985. Direct evidence for chromosomal inversion during T-cell receptor β gene rearrangements. *Nature*, in press.

87. Patten, P., Yokota, T., Rothbard, J., Chien, Y.-H., Arai, K.-I., Davis, M. M. 1984. Structure, expression and divergence of T-cell receptor β-chain variable regions. *Nature* 312:40-46.

88. Barth, R., Kim, B., Lan, N., Hunkapiller, T., Sobieck, N., Winoto, A., Gershenfeld, H., Okada, C., Hansburg, D., Weissman, I., Hood, L. 1985. The murine T-cell receptor employs a limited repertoire of expressed V_β gene segments. *Nature* 316:517-523.

89. Behlke, M. A., Spinella, D. G., Chou, H., Sha, W., Hartl, D. L., Loh, D. Y. 1985. T-cell receptor β-chain expression: Dependence on relatively few variable region genes. *Science* 229:566-570.

90. Cory, S., Tyler, B., Adams, J. 1981. Sets of immunoglobulin V_κ genes homologous to 10 cloned V_κ sequences: implications for the number of germline V_κ genes. *J. Mol. Applied Genet.* 1:103-116.

91. Brodeur, P., Riblet, R. 1984. The immunoglobulin heavy chain variable region (Igh-V) locus in mouse. I. One hundred Igh-V genes comprise seven families of homologous genes. *Eur. J. Immunol.* 14:922-930.

92. Dildrop, R. 1984. A new classification of mouse V_H sequences. *Immunol. Today* 5:85-86.

93. Siu, G., Strauss, E., Hood, L. Analysis of a human V_β gene subfamily. Submitted for publication.

94. Arden, B., Klotz, J., Siu, G., Hood, L. 1985. Diversity and structure of genes of the α family of mouse T-cell antigen receptor. *Nature* 316:783-787.

95. Becker, D., Patten, P., Chien, Y.-H., Yokota, T., Eshhar, Z., Giedlin, M., Gascoigne, N. R. J., Goodenow, C., Wolf, R., Arai, K.-I., Davis, M. M. Variability and repertoire size in T-cell receptor V_α gene segments. *Nature*, in press.

96. Sims, J. E., Tunnacliffe, A., Smith, W. G., Rabbitts, T. H. 1984. Complexity of human T-cell antigen receptor beta-chain constant- and variable-region genes. *Nature* 312:541-545.

97. Yoshikai, Y., Anatoniou, D., Clark, S. P., Yanagi, Y., Sangster, R., van den Elsen, P., Terhorst, C., Mak, T. W. 1984. Sequence and expression of transcripts of the human T-cell receptor β-chain genes. *Nature* 312:521-524.

98. Jones, N., Leiden, J., Dialynas, D., Fraser, J., Clabby, M., Kishimoto, T., Strominger, J. L., Andrews, D., Lane, W., Woody, J. 1985. Partial primary structure of the alpha and beta chains of human tumor T-cell receptors. *Science* 227:311-314.

99. Tunnacliffe, A., Kefford, R., Milstein, C., Forster, A., Rabbitts, T. H. 1985. Sequence and evolution of the human T-cell receptor β-chain genes. *Proc. Natl. Acad. Sci.* 82:5068-5072.

100. Honjo, T., Obata, M., Yamawaki-Katoaka, Y., Kataoka, T., Takahashi, N., Mano, Y. 1979. Cloning and complete nucleotide sequence of mouse immunoglobulin γ1 chain gene. *Cell* 18:559-568.

101. Tucker, P., Marcu, K., Newell, N., Richard, J., Blattner, F. 1979. Sequence of the cloned gene for the constant region of the murine γ2b immunoglobulin heavy chain. *Science* 206:1303-1306.

102. Cheng, H.-L., Blattner, F., Fitzmaurice, L., Mushinski, J., Tucker, P. 1982. Structure of genes for membrane and secreted murine IgD heavy chains. *Nature* 296:410-415.

103. Ellison, J., Berson, B., Hood, L. 1982. The nucleotide sequence of a human $C_{\gamma 1}$ gene. *Nucl. Acids. Res.* 10:4071-4079.

104. Ellison, J., Hood, L. 1982. Linkage and sequence homology of two human immunoglobulin γ heavy chain constant region genes. *Proc. Natl. Acad. Sci. USA* 79:1984-1988.

105. Ishida, N., Ueda, S., Hayashida, H., Miyata, T., Honjo, T. 1982. The nucleotide sequence of the mouse immunoglobulin epsilon gene: comparison with the human epsilon gene sequence. *EMBO J.* 1:1117-1123.

106. Lefranc, M. P., Rabbitts, T. H. 1985. Two tandemly organised human genes encoding the T-cell γ constant region sequences show multiple rearrangement in different T-cell lines. *Nature* 316:464-466.

107. Hood, L., Campbell, J., Elgin, S. R. C. 1975. The organization, expression, and evolution of antibody genes and other multigene families. *Ann. Rev. Genetics* 9:305-354.

108. Williams, A. R., Gagnon, J. 1982. Neuronal cell Thy-1 glycoprotein: homology with immunoglobulin. *Science* 216:696-703.

109. Hood, L., Kronenberg, M., Hunkapiller, T. 1985. T-cell antigen receptors and the immunoglobulin supergene family. *Cell* 40:225-229.

110. Strominger, J., Orr, T., Pachan, P., Ploegh, H. L., Mann, D. L., Bilofsky, H., Saroff, H. A., Wu, T. T., Kabat, E. A. 1980. An evaluation of the significance of amino acid sequence homologies in human histocompatibility antigens (HLA-A and HLA-B). *Scand. J. Immunol.* 11:573-593.

111. Steinmetz, M., Frelinger, J. G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S. M., Uehara, H., Nathanson, S., Hood, L. 1981. Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. *Cell* 24:125-134.

112. Larhammar, D., Gustafsson, K., Claesson, L., Bill, P., Wiman, K., Schenning, L., Sundelin, J., Widmark, E., Peterson, P. A., Rask, L. 1982. Alpha chain of HLA-DR transplantation antigens is a member of the same protein superfamily as the immunoglobulins. *Cell* 30:153-161.

113. Malissen, M., Hunkapiller, T., Hood, L. 1983. Nucleotide sequence of a light chain gene of the mouse I-A subregion: A_{β}^{d}. *Science* 221:750-754.

114. Campbell, D. G., Williams, A. F., Bayley, P. M., Reid, K. B. M. 1979. Structural similarities between thy-1 antigen from rat brain and immunoglobulin. *Nature* 282:341-343.

115. Clark, M. J., Gagnon, J., Williams, A. F., Barclay, A. N. 1985. MRC OX-2 antigen: a lymphoid/neuronal membrane glycoprotein with a structure like a single immunoglobulin light chain. *EMBO J.* 4:113-118.

116. Mostov, K. E., Friedlander, M., Blobel, G. 1984. The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. *Nature* 308:37-43.

116a. Maddon, P. J., Littman, D. R., Godfrey, M., Maddon, D. E., Chess, L., Axel, R. 1985. The isolation and nucleotide sequence of a cDNA encoding the T-cell surface protein T4: A new member of the immunoglobulin gene family. *Cell* 42:93-104.

117. Littman, D. R., Thomas, Y., Maddon, P. J., Chess, L., Axel, R. 1985. The isolation and sequence of the gene encoding T8: A molecule defining functional classes of T lymphocytes. *Cell* 40:237-246.

118. Sukhatme, V. P., Sizer, K. C., Vollmer, A. C., Hunkapiller, T., Parnes, J. R. 1985. The T cell differentiation antigen Leu-2/T8 is homologous to immunoglobulin and T cell receptor variable regions. *Cell* 40:591-597.

119. Peterson, P. A., Cunningham, B. A., Berggard, I., Edelman, G. M. 1972. β_2-microglobulin - a free immunoglobulin domain. *Proc. Natl. Acad. Sci. USA* 69:1697-1701.

120. Amzel, L. M., Poljak, R. J. 1979. Three-dimensional structure of immunoglobulins. *Ann. Rev. Biochem.* 48:961-997.

121. Huang, H., Hood, L. 1982. The ontogenetic and evolutionary origins of antibody diversity. In: *Stadler Symposium* 14:59-68.

122. Cory, S., Adams, J. M. Kemp, D. J. 1980. Somatic rearrangements forming active immunoglobulin μ genes in B and T lymphoid cell lines. *Proc. Natl. Acad. Sci. USA* 77:4943-4947.

123. Forster, A., Hobart, M., Hengartner, H., Rabbitts, T. H. 1980. An immunoglobulin heavy-chain gene is altered in two T-cell clones. *Nature* 286:897-899.

124. Kurosawa, Y., von Boehmer, H., Haas, W., Sakano, H., Trauneker, A., Tonegawa, S. 1981. Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. *Nature* 290:566-570.

125. Kronenberg, M., Kraig, E., Horvath, S. J., Hood, L. E. 1982. Cloned T cells as a tool for molecular geneticists: Approaches to cloning genes which encode T-cell antigen receptors. In: *Isolation, Characterization and Utilization of T Lymphocyte Clones.* (eds C. G. Fathman and F. Fitch) Academic Press, N. Y. 467-491.

126. Zuñiga, M. C., D'Eustachio, P., Ruddle, N. 1982. Immunoglobulin heavy chain gene rearrangment and transcription in murine T cell hybrids and T lymphomas. *Proc. Natl. Acad. Sci. USA* 79:3015-3019.

127. Kemp, D. J., Adams, J. M., Mottram, P. L., Thomas, W. R., Walker, I. D., Miller, J. F. A. P. 1982. A search for messenger RNA molecules bearing immunoglobulin V_H nucleotide sequences in T cells. *J. Exp. Med.* 156:1848-1853.

128. Nakanishi, K., Sugimura, K., Yaoita, Y., Maeda, K., Kashiwamura, S.-I., Honjo, T., Kishimoto, T. 1982. A T15-idiotype-positive T suppressor hybridoma does not use the T15 V_H gene segment. *Proc. Natl. Acad. Sci. USA* 79:6984-6988.

129. Kraig, E., Kronenberg, M., Kapp, J., Pierce, C. W., Abruzzini, A. F., Sorensen, C. M., Samelson, L. E., Schwartz, R. H., Hood, L. E. 1983. T and B cells that recognize the same antigen do not transcribe similar heavy chain variable regions gene segments. *J. Exp. Med.* 158:192-209.

130. Kronenberg, M., Kraig, E., Siu, G., Kapp, J. A., Kappler, J., Marrack, P., Pierce, C. W., Hood, L. 1983. Three T-cell hybridomas do not express detectable heavy chain variable gene transcripts. *J. Exp. Med.* 158:210-227.

131. Kronenberg, M., Davis, M. M., Early, P. W., Hood, L. E., Watson, J. D. 1980. Helper and killer T cells do not express B cell immunoglobulin joining and constant region gene segments. *J. Exp. Med.* 152:1745-1761.

132. Flug, F., Pelicci, P.-G., Bonetti, F., Knowles, D. M. II., Dalla-Favera, R., 1985. T-cell receptor gene rearrangements as markers of lineage and clonality in T-cell neoplasms. *Proc. Natl. Acad. Sci. USA* 82:3460-3464.

133. O'Connor, N. T. J., Wainscoat, J. S., Weatherall, D. J., Gatter, K. C., Feller, A. C., Isaacson, P., Jones, D., Lennert, K., Pallesen, G., Ramsey, A., Stein, H., Wright, D. H., Mason, D. Y. 1985. Rearrangement of the T-cell receptor β-chain gene in the diagnosis of lympho-proliferative disorders. *Lancet* 1:1295-1297.

134. Rabbitts, T. H., Stinson, A., Forster, A., Foconi, L., Luzzatoo, L., Catovsky, D., Hammarström, L., Smith, C. I. E., Jones, D., Karpas, A., Minowada, J., Taylor, A. M. R. 1985. Heterogeneity of T-cell β-chain gene rearrangements in human leukaemias and lymphomas. *EMBO J.* 4:2217-2224.

135. Alt, F. W., Yancopoulos, G. D., Blackwell, T. K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., Baltimore, D., 1984. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J.* 3:1209-1219.

136. Steinmetz, M., Altenburger, W., Zachau, H. G. 1980. A rearranged DNA sequence possibly related to the translocation of immunoglobulin gene segments. *Nucl. Acids. Res.* 8:1709-1720.

137. Selsing, E., Storb, U. 1981. Mapping of immunoglobulin variable region genes: relationship to the "deletion" model of immunoglobulin gene rearrangement. *Nucl. Acids. Res.* 9:5725-5735.

138. Höchtl, J., Müller, C. R., Zachau, H. G. 1982. Recombined flanks of the variable and joining segments of immunoglobulin genes. *Proc. Natl. Acad. Sci.* 79:1383-1387.

139. Lewis, S., Rosenberg, N., Alt, F., Baltimore, D. 1982. Continuing kappa-gene rearrangement in a cell line transformed by Abelson Murine Leukemia Virus. *Cell* 30:807-816.

140. Van Ness, B. G., Coleclough, C., Perry, R. P., Weigert, M. 1982. DNA between variable and joining gene segments of immunoglobulin κ light chain is frequently retained in cells that rearrange the κ locus. *Proc. Natl. Acad. Sci. USA* 79:262-266.

141. Höchtl, J., Zachau, H. G. 1983. A novel type of aberrant recombination in immunoglobulin genes and its implications for V-J joining mechanisms, *Nature* 302:260-263.

142. Hozumi, N., Tonegawa, S. 1976. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. USA* 73:3628-3232.

143. Lewis, S., Gifford, A., Baltimore, D. 1984. Joining of V_κ to J_κ gene segments in a retroviral vector introduced into lymphoid cells. *Nature* 308:425-428.

144. Kronenberg, M., Goverman, J., Haars, R., Malissen, M., Kraig, E., Phillips, L., Delovitch, T., Suciu-Foca, N., Hood, L. 1985. Rearrangement and transcription of the β-chain genes of the T-cell antigen receptor in different types of murine lymphocytes. *Nature* 313:647-653.

145. Duby, A. D., Klein, K. A., Murre, C., Seidman, J. G. 1985. A novel mechanism of somatic rearrangement predicted by a human T-cell antigen receptor β-chain complementary DNA. *Science* 228:1204-1206.

146. Haars, R., Kronenberg, M., Owen, F., Gallatin, M., Weissman, I., Hood, L. Rearrangement and expression of T-cell antigen receptor and γ chain genes during thymic differentiation. Submitted for publication.

147. Kronenberg, M., Haars, R., Gartland, L., Hood, L., Cooper, M. Rearrangement of T-cell receptor beta genes in human peripheral blood cells that express surface markers characteristic of NK cells. Submitted for publication.

148. Goverman, J., Minard, K., Shastri, N., Hunkapiller, T., Hansburg, D., Sercarz, E., Hood, L. 1985. Rearranged β T-cell receptor genes in a helper T cell clone specific for lysozyme: no correlation between V_β and MHC restriction. *Cell* 40:859-867.

149. Kranz, D. M., Saito, H., Heller, M., Takagaki, Y., Haas, W., Eisen, H. N., Tonegawa, S. 1985. Limited diversity of the rearranged T-cell γ gene. *Nature* 313:752-755.

150. Royer, H. D., Bensussan, A., Acuto, O., Reinherz, E. L. 1984. Functional isotypes are not encoded by the constant region genes of the β subunit of the T-cell receptor for antigen/major histocompatibility complex. *J. Exp. Med.* 160:947-952.

151. Toyonaga, B., Yanagi, Y., Suciu-Foca, N., Minden, M., Mak, T. W. 1984. Rearrangements of T-cell receptor gene YT35 in human DNA from thymic leukaemia T-cell lines and functional T-cell clones. *Nature* 311:385-387.

152. Minden, M. P., Toyonaga, B., Ha, K., Yanagi, Y., Chin, B., Gelfand, E., Mak, T. 1985. Somatic rearrangment of T-cell antigen receptor gene in human T-cell malignancies. *Proc. Natl. Acad. Sci. USA* 82:1224-1227.

153. Hedrick, S. M., Germain, R. N., Bevan, M. J., Dorf, M., Engel, I., Fink, P., Gascoigne, N., Heber-Katz, E., Kapp, J., Kaufmann, Y., Kaye, J., Melchers, F., Pierce, C., Schwartz, R. H., Sorensen, C., Taniguchi, M., Davis, M. M. 1985. Rearrangment and transcription of a T-cell receptor β-chain gene in different T-cell subsets. *Proc. Natl. Acad. Sci. USA* 82:531-535.

154. Bensussan, A., Acuto, O., Hussey, R. E., Milanese, C., Reinherz, E. L. 1984. T3-Ti receptor triggering of T8+ suppressor T cells leads to unresponsiveness to interleukin-2. *Nature* 311:565-567.

155. Yanagi, Y., Caccia, N., Kronenberg, M., Chin, B., Roder, J., Rohel, D., Kiyohare, T., Lauzon, R., Toyonaga, B., Rosenthal, G., Dennert, H., Acha-Orbea, H., Hengartner, H., Hood, L., Mak, T. W. 1985. Gene rearrangement in cells with natural killer activity and expression of the ß-chain of the T-cell antigen receptor. *Nature* 314:631-633.

156. Reynolds, C. W., Bonyhadi, M., Herberman, R. B., Young, H. A., Hedrick S. M. 1985. Lack of gene rearrangement and mRNA expression of the beta-chain of the T cell receptor in spontaneous rat large granular lymphocyte leukemia lines. *J. Exp. Med.* 161:1249-1254.

157. Ritz, J., Campen, T. J., Schmidt, R. E., Royer, H. D., Hercend, T., Hussey, R. E., Reinherz, E. L. 1985. Analysis of T-cell receptor gene rearrangement and expression in human natural killer clones. *Science* 228:1540-1543.

158. Reth, M. G., Alt, F. W. 1984. Novel immunoglobulin heavy chains are produced from DJ_H gene segment rearrangements in lymphoid cells. *Nature* 312:418-423.

159. Rupp, F., Acha-Orbea, H., Hengartner, H., Zinkernagel, R., Joho, R. 1985.. Identical V_β T-cell receptor genes used in alloreactive cytotoxic and antigen plus I-A specific helper T cells. *Nature* 315:425-427.

160. Raulet, D. H., Garman, R. D., Saito, H., Tonegawa, S. 1985. Developmental regulation of T-cell receptor gene expression. *Nature* 314:103-107.

161. Honjo, T. 1983. Immunoglobulin genes. *Ann. Rev. Immunol.* 1:499-528.

162. Kotzin, B. L., Barr, V., Palmer, E. 1985. A large deletion within the T-cell receptor beta chain gene complex in NZW mice. *Science* 229:167-171.

163. Alt, F. W. 1984. Exclusive Immunoglobulin genes. *Nature* 312:502-503.

164. Sugiyama, H., Akira, S., Kikutani, H., Kishimoto, S., Yamamura, Y., Kishimoto, T. 1983. Functional V region formation during *in vitro* culture of murine immature B precursor cell line. *Nature* 303:812-815.

165. Alt. F. W., Rosenberg, N., Enea, V., Siden, E., Baltimore, D. 1982. Multiple immunoglobulin heavy-chain gene transcripts in Abelson murine leukemia virus-transformed lymphoid cell lines. *Mol. and Cell Biol.* 2:386-400.

166. Alt, F., Rosenberg, N., Lewis, S., Thomas, E., Baltimore, D. 1981. Organization and reorganization of immunoglobulin genes in A-MULV transformed cells: rearrangment of heavy but not light chain genes. *Cell* 27:381-390.

167. Alt, F. W., Enea, V., Bothwell, A. L. M., Baltimore, D. 1980. Activity of multiple light chain genes in murine myeloma cells producing a single, functional light chain. *Cell* 21:1-12.

168. Kwan, S.-P., Max, E. E., Seidman, J. G., Leder, P., Scharff, M. D. 1981. Two kappa immunoglobulin genes are expressed in the myeloma S107. *Cell* 26:57-66.

169. Hieter, P. A., Korsmeyer, S. J., Waldmann, T. A., Leder, P. 1981. Human immunoglobulin κ light-chain genes are deleted or rearranged in λ-producing B cells. *Nature* 290:368-372.

170. Widmer, M. D., MacDonald, H. R., Cerottini, J.-C. 1981. Limiting dilution analysis of alloantigen-reactive T lymphocytes VI. Ontogeny of cytolytic T lymphocyte precursors in the thymus. *Thymus* 2:245-255.

171. Moore, M. A. S., Owen, J. J. T. 1967. Experimental studies on the development of the thymus. *J. Exp. Med.* 126:715-725.

172. Snodgrass, H. R., Kisielow, P., Kiefer, M., Steinmetz, M., von Boehmer, H. 1985. Ontogeny of the T-cell antigen receptor within the thymus. *Nature* 313:592-595.

173. Snodgrass, H. R., Dembic, Z. Steinmetz, M., von Boehmer, H. 1985. Expression of T-cell antigen receptor genes during fetal development in the thymus. *Nature* 315:232-233.

174. van Ewijk, W., Jenkinson, E, J., Owen, J. J. T. 1982. Detection of Thy-1, T-200, Lyt-1, and Lyt-2 bearing cells in the developing lymphoid organs of the mouse embryo *in vivo* and *in vitro*. *Eur. J. Immunol.* 12:262-271.

175. Ceredig, R., MacDonald, H. R., Jenkinson, E. J. 1983. Flow microfluorometric analysis of mouse thymus development *in vivo* and *in vitro*. *Eur. J. Immunol.* 13:185-190.

176. Habu, S., Okumura, K. 1984. Cell surface antigen marking the stages of murine T cell ontogeny and its functional subsets. *Immunol. Rev.* 82:117-139.

177. Takacs, L., Osawa, H., Diamantstein, T. 1984. Detection and localization by the monoclonal anti-interleukin-2 receptor antibody AMT-13 of IL2 receptor-bearing cells in the developing thymus of the mouse embryo and in the thymus of cortisone-treated mice. *Eur. J. Immunol.* 14:1152-1156.

178. Habu, S., Okumura, K., Diamantstein, T., Shevach, E. M. 1985. Expression of interleukin 2 receptor on murine fetal thymocytes. *Eur. J. Immunol.* 15:456-460.

179. Roehm, N., Herron, L., Cambier, J., DiGuisto, D., Haskins, K., Kappler, J., Marrack, P. 1984. The major histocompatibility complex-restricted antigen receptor on T cells: distribution on thymus and peripheral T cells. *Cell* 38:577-584.

180. Born, W., Yagüe, J., Palmer, E., Kappler, J., Marrack, P. 1985. Rearrangement of T-cell receptor β-chain genes during T-cell development. *Proc. Natl. Acad. Sci. USA* 82:2925-2929.

181. Samelson, L. E., Lindsten, T., Fowlkes, B. J., van den Elsen, P., Terhorst, C., Davis, M. M., Germain, R. N., Schwartz, R. H. 1985. Expression of genes of the T-cell antigen receptor complex in precursor thymocytes. *Nature* 315:765-768.

182. Trowbridge, I. S., Lesley, J., Trotter, J., Hyman, R. 1985. Thymocyte subpopulation enriched for progenitors with an unrearranged T-cell receptor β-chain gene. *Nature* 315:666-669.

183. Fowlkes, B. J., Edison, L., Mathieson, B., Chused, T. 1984. Differentiation *in vitro* of an adult precursor thymocyte. In *Regulation of the Immune System UCLA Symposia on Molecular and Cellular Biology.*, New Series, Vol 18. eds. Cantor, H., Chess, L., and Sercarz, E. Alan R. Liss, Inc., New York, 1984, pp. 275-293.

184. Mathieson, B. J., Fowlkes, B. J. 1984. Cell surface antigen expression on thymocytes: development and phenotypic differentiation of intrathymic subsets. *Immunol. Rev.* 82:141-173.

185. Kinnon, C., Diamond, R., Rothenberg, E. Expression of T-cell antigen receptor mRNAs in the thymus: implications for cell lineages. Submitted for publication.

186. Shortman, K., Jackson, H. 1974. The differentiation of T lymphocytes. I. Proliferation kinetics and interrelationships of subpopulations of mouse thymus cells. *Cell. Immunol.* 12:230-246.

187. Robinson, J. H., Owen, J. J. T. 1976. Generation of T-cell function in organ culture of foetal mouse thymus. *Clin. Exp. Immunol.* 23:347-354.

188. Robinson, J. H., Owen, J. J. T. 1977. Generation of T-cell function in organ culture of foetal mouse thymus. II. Mixed lymphocyte culture reactivity. *Clin. Exp. Immunol.* 27:322-327.

189. Ceredig, R., Jenkinson, E. J., MacDonald, H. R., Owen, J. J. T. 1982. Development of cytolytic T lymphocyte precursors in organ-cultured mouse embryonic thymus rudiments. *J. Exp. Med.* 155:617-622.

190. Owen, J. J. T., Jenkinson, E. J., Kingston, R. 1983. The ontogeny of T lymphocytes. *Ann. Immunol. (Inst. Pasteur)* 134D:115-122.

191. Kisielow, P., Leiserson, W., von Boehmer, H. 1984. Differentiation of thymocytes in fetal organ culture: analysis of phenotypic changes accompanying the appearance of cytolytic and interleukin 2-producing cells. *J. Immunol.* 133:1117-1123.

192. Teh, H.-S., Ho, M. 1985. Ontogeny of proliferative and cytotoxic responses to interleukin 2 and concanavalin A in murine fetal thymus. *J. Immunol.* 134:1653-1658.

193. Royer, H. D., Acuto, O., Fabbi, M., Tizard. R., Ramachandran, K., Smart, J. E., Reinherz, E. L. 1984. Genes encoding the Tiß subunit of the antigen/MHC receptor undergo rearrangement during intrathymic ontogeny prior to surface T3-Ti expression. *Cell* 39:261-266.

194. Acuto, O., Campen, T. J., Royer, H. D., Hussey, R. E., Poole, C. B., and Reinherz, E. L. 1985. Molecular analysis of T cell receptor (Ti) variable region (V) gene expression. Evidence that a single Tiß V gene family can be used in formation of V domains on phenotypically and functionally diverse T cell populations. *J. Exp. Med.* 161:1326-1343.

195. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W., Tonegawa, S. 1980. Two types of somatic recombination are necessary for generation of complete immunoglobulin heavy chain genes. *Nature* 286:676-683.

196. Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J., Hood, L. 1980. The joining of V and J gene segments creates antibody diversity. *Nature* 283:497-499.

197. Alt. F., Baltimore, D. 1982. Joining of immunoglobulin heavy chain gene segments: implications from a chromosome with evidence of three D-J_H fusions. *Proc. Natl. Acad. Sci. USA* 78:5812-5816.

198. Weigert, M. G., Cesari, I. M., Yonkovich, S. J., Cohn, M. 1970. Variability in the lambda light chain sequences of mouse antibody. *Nature* 228:1045-1047.

199. Kim, S., Davis, M., Sinn, E., Patten, P., Hood, L. 1981. Antibody diversity: somatic hypermutation of rearranged V_H genes. *Cell* 27:573-581.

200. Pech, M., Höchtl, J., Schnell, H., Zachau, H. G. 1981. Differences between germ-line and rearranged immunoglobulin V_κ coding sequences suggest a localized mutation mechanism. *Nature* 291:668-670.

201. Bauer, D. C., Mathies, M. J., Stavitsky, A. A. 1963. Sequences of synthesis of γ-1 macroglobulin and γ-2 globulin antibodies during primary and secondary responses to proteins, Salmonella antigens, and phage. *J. Exp. Med.* 117:889-907.

202. Augustin, A. A., Sim, G. K. 1984. T-cell receptors generated via mutations are specific for various major histocomptibility antigens. *Cell* 39:5-12.

203. Schwartz, R. H. 1985. T lymphocyte recognition of antigen associated with the gene product of the major histocompatibility complex. *Ann. Rev. Immunol.* 3:237-261.

204. Cohen, R. J., Eisen, H. 1977. Interactions of macromolecules on cell menbranes and restrictions of T-cell specificity by products of the major histocompatibility complex. *Cell. Immunol.* 32:1-9.

205. Matzinger, P. 1981. A one-receptor view of T-cell behavior. *Nature* 292:497-501.

206. Schrader, J. 1982. A single T-cell receptor: A speculative review of the intrathymic generation and modulation of the repertoire. *Thymus* 4:181-207.

207. Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E., Unanue, E. R. 1985. The binding of immunogenic peptides to Ia histocompatibility molecules. *Nature* 317:359-361.

208. Blanden, R. V., Ada, G. L. 1978. A dual recognition model for cytotoxic T cells based on thymic selection of precursors with low affinity for self H-2 antigens. *Scand. J. Immunol.* 7:181-190.

209. Cohn, M., Epstein, R. 1978. T-cell inhibition of humoral responsiveness. II. Theory on the role of restrictive recognition in immune regulation. *Cell. Immunol.* 39:125-153.

210. von Boehmer, H., Haas, W., Jerne, N. K. 1978. Major histocompatibility complex-linked immune-responsiveness is acquired by lymphocytes of low-responder mice differentiating in thymus of high-responder mice. *Proc. Natl. Acad. Sci. USA* 75:2439-2442.

211. Parham, P. 1984. A repulsive view of MHC-restriction. *Immunol. Today* 5:89-92.

212. Pernis, B., Axel, R. 1985. A one and a half receptor model for MHC-restricted antigen recognition by T lymphocytes. *Cell* 41:13-16.

213. Cleveland, W. L., Elanger, B. F. Hypothesis: The MHC-restricted T-cell receptor as a structure with two multistate allosteric combining sites. *Molec. Immunol.*, in press.

214. Kappler, J. W., Skidmore, B., White, J., Marrack, P. 1981. Antigen-inducible, H-2-restricted interleukin-2-producing T cell hybridomas. Lack of independent antigen and H-2 recognition. *J. Exp. Med.* 153:1198-1214.

215. Marrack, P., Shimonkevitz, R., Hannum, C., Haskins, K., Kappler, J. 1983. The major histocompatibility complex-restricted antigen receptor on T cells. IV. An anti-idiotypic antibody predicts both antigen and I-specificity. *J. Exp. Med.* 158:1635-1646.

216. Hünig, T. R., Bevan, M. J., 1982. Antigen recognition by cloned cytotoxic T-lymphocytes follow rules predicted by the altered self hypothesis. *J. Exp. Med.* 155:111-125.

217. Heber-Katz, E., Schwartz, R. H., Matis, L. A., Hannum, C., Fairwell, T., Appella, E., Hansburg, D. 1982. Contribution of antigen-presenting cell major histocompatibility complex gene products to the specificity of antigen-induced T-cell activation. *J. Exp. Med.* 155:1086-1099.

218. Hedrick, S. M., Matis, L. A., Hecht, T. T., Samelson, L. E., Longo, D. L., Heber-Katz, E., Schwartz, R. H. 1982. The fine specificity of antigen and Ia determinant recognition by T cell hybridoma clones specific for cytochrome c. *Cell* 30:141-152.

219. Matis, L. A., Longo, D. L., Hedrick, S. M., Hannum, C., Margoliash, E., Schwartz, R. H. 1983. Clonal analysis of the major histocompatibility complex restriction and the fine specificity of antigen recognition in the T cell proliferative response to cytochrome c. *J. Immunol.* 130:1527-1535.

220. Langman, R. E., Cohn, M. T-cells function via restricted recognition of antigen, not antigen-restricted recognition. *Cell. Immunol.*, in press.

221. Groves, E. S., Singer, A. 1983. Role of the H-2 complex in the induction of T cell tolerance to self minor histocompatibility antigens. *J. Exp. Med.* 158:1483-1497.

222. Matzinger, P., Zamoyska, R., Waldmann, H. 1984. Self tolerance is H-2 restricted. *Nature* 308:738-741.

223. Rammensee, H.-G., Bevan, M. J. 1984. Evidence from *in vitro* studies that tolerance to self antigens is MHC-restricted. *Nature* 308:741-744.

224. Benacerraf, B. 1978. A hypothesis to relate the specificity of T lymphocytes and the activity of I region-specific Ir genes in macrophages and B lymphocytes. *J. Immunol.* 120:1809-1812.

225. Rosenthal, A. S. 1978. Determinant selection and macrophage function in genetic control of the immune response. *Immunol. Rev.* 40:136.

226. Davies, D. R., Metzger, H. 1983. Structural basis of antibody function. *Ann. Rev. Immunol.* 1:87-117.

227. Kyte, J., Doolittle, R. G. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105-132.

228. Chou, P. Y., Fassman, G. D. 1978. Empirical predictions of protein conformation. *Ann. Rev. Biochem.* 47:251-276.

229. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M., Perry, H. 1983. Sequences of immunological interest. U. S. Department of Health Services, Washington, D. C.

230. Segal, D., Padlan, E., Cohen, G., Rudikoff, S., Potter, M., Davies, D. 1974. The three-dimensional structure of a phosphorylcholine-binding mouse immunoglobulin Fab and the nature of the antigen binding site. *Proc. Natl. Acad. Sci. USA* 71:4298-4302.

231. Saul, F., Amzel, L., Poljak, R. 1978. Preliminary refinement and structural analysis of the Fab fragment from human immunoglobulin New at 2.0 Å resolution. *J. Biol. Chem.* 253:585-597.

232. Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M., Westholm, F. A., Fausch, M. D., Deutsch, H. F. 1974. Binding of 2, 4-dinitrophenyl compounds and other small molecules to a crystalline λ-type Bence-Jones dimer. *Biochemistry* 13:3816-3827.

233. Wu, T., Kabat, E. 1970. Analysis of the sequences of Bence-Jones proteins and myeloma light chains and their implications of antibody complementarity. *J. Exp. Med.* 132:211-250.

234. Capra, J. D., Wasserman, R. L., Kehoe, J. M. 1973. Phylogenetically associated residues within the V_{HIII} subgroup of several mammalian species. Evidence for a "pauci-gene" basis for antibody diversity. *J. Exp. Med.* 138:410-427.

235. Loh, E., Black, B., Riblet, R., Weigert, M., Hood, J. M., Hood, L. 1979. Myeloma proteins from NZB and BALB/c mice: structural and functional differences. *Proc. Natl. Acad. Sci. USA* 76:1395-1399.

236. Loh, E., Hood, J. M., Riblet, R., Weigert, M., Hood. L. 1979. Comparisons of myeloma proteins from NZB and BALB/c mice: structural and functional differences of heavy chains. *J. Immunol.* 122:44-48.

237. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V., Poljak, R. J., 1985. Three-dimensional structure of an antigen-antibody complex at 6 Å resolution. *Nature* 313:156-158.

238. Wylie, D. E., Sherman, L. A., Klinman, N. R. 1982. Participation of the major histocompatibility complex in antibody recognition of viral antigens expressed on infected cells. *J. Exp. Med.* 155:403-414.

239. Norcross, M. A., Kanehisa, M. 1985. The predicted structure of the Ia β_1 domain: A hypothesis for the structural basis of MHC-restricted T-cell recognition of anitigens. *Scand. J. Immunol.* 21:511-523.

240. Clarke, S. H., Huppi, K., Ruezinsky, D., Staudt, L., Gerhard, W., Weigert, M. 1985. Inter- and intraclonal diversity in the antibody response to influenza hemagglutinin. *J. Exp. Med.* 161:687-707.

241. Crews, S., Griffin, J., Huang, H., Calame, K., Hood, L. 1981. A single V_H gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of antibody. *Cell* 25:59-66.

242. Alter, B. J., Schendel, D. J., Bach, M. L., Bach, F. H. Klein, J., Stimpfling, J. H. 1973. Cell-mediated lympholysis. Importance of serologically defined H-2 regions. *J. Exp. Med.* 137:1303-1309.

243. Zinkernagel, R. M., Doherty, P. C. 1975. H-2 compatibility requirement for T-cell-mediated lysis of target cells infected with lymphocytic choriomeningitis virus. *J. Exp. Med.* 141:1427-1436.

244. Zinkernagel, R. M., Doherty, P. C. 1979. MHC restricted cytotoxic T cells: Studies on the biological role of polymorphic major transplantation antigens determining T cell restriction specificity, function and responsiveness. *Adv. Immunol.* 51-177.

245. Meo, T., David, C. S., Rijnbeek, H. M., Nabholz, M., Miggiano, V. C., Shreffler, D. C. 1975. Inhibition of mouse MLR by anti-Ia sera. *Transplantation Proceedings.* 7:127-129.

246. Katz, D. H., Graves, M., Dorf, M. E., Dimuzio, H., Benacerraf, B. 1975. Cell interactions between histoincompatible T and B lymphocytes. VII. Cooperative responses between lymphocytes are controlled by genes in the I region of the H-2 complex. *J. Exp. Med.* 141:263-268.

247. Thomas, D. W., Yamashita, U., Shevach, E. M. 1977. Nature of the antigenic complex recognized by T lymphycytes. IV. Inhibition of antigen-specific proliferation by antibodies to stimulator macrophage Ia antigens. *J. Immunol.* 119:223-226.

248. Swierkosz, J. E., Marrack, P., Kappler, J. W. 1979. The role of H-2 linked genes in helper T cell function. V. I-region control of helper T cell interaction with antigen-presenting macrophages. *J. Immunol.* 123:654-659.

249. Zinkernagel, R. M., Callahan, G. N., Althage, A., Cooper, S., Klein, P. A., Klein, J. 1978. On the thymus in the differentiation of "H-2 self-recognition" by T cells: evidence for dual recognition? *J. Exp. Med.* 147:882-896.

250. Klein, J., Nagy, Z. A. 1982. MHC restriction and Ir genes. *Adv. Cancer Res.* 37:234-317.

251. Schwartz, R. H. 1984. The role of gene products of the major histocompatibility complex in T-cell activation and cellular interactions. In: *Fundamental Immunology*, ed. W. E. Paul, Raven Press, New York, pp. 379-438.

252. Bryant, B. J. 1972. Renewal and fate in the mammalian thymus: mechanisms and inferences of thymocytokinetics. *Eur. J. Immunol.* 2:38-45.

253. McPhee, D., Pye, J., Shortman, K. 1979. The differentiaton of T lymphocytes. V. Evidence for intrathymic death of most thymocytes. *Thymus* 1:151-162.

254. Scollay, R., Butcher, E., Weissman, I. 1980. Thymus migration: Quantitative studies on the rate of migration of cells from the thymus to the periphery in mice. *Eur. J. Immunol.* 10:210-218.

255. Paul, W. E. 1984. Immune response genes. In: *Fundamental Immunology*, ed. W. E. Paul, Raven Press, New York.

256. Longo, D. L., Matis, L. A., Schwartz, R. H. 1981. Insights into immune response gene function from experiments with chimeric animals. *Crit. Rev. Immunol.* 2:83-132.

257. Zinkernagel, R. M., Althage, A., Cooper, S., Kreeb, G., Klein, P. A., Sefton, B. Flaherty, L. Stimpfling, J., Shreffler, D., Klein, J. 1978. I-r genes in H-2 regulate generation of anti-viral cytotoxic T cells. Mapping to K or D and dominance of unresponsiveness. *J. Exp. Med.* 148:592-606.

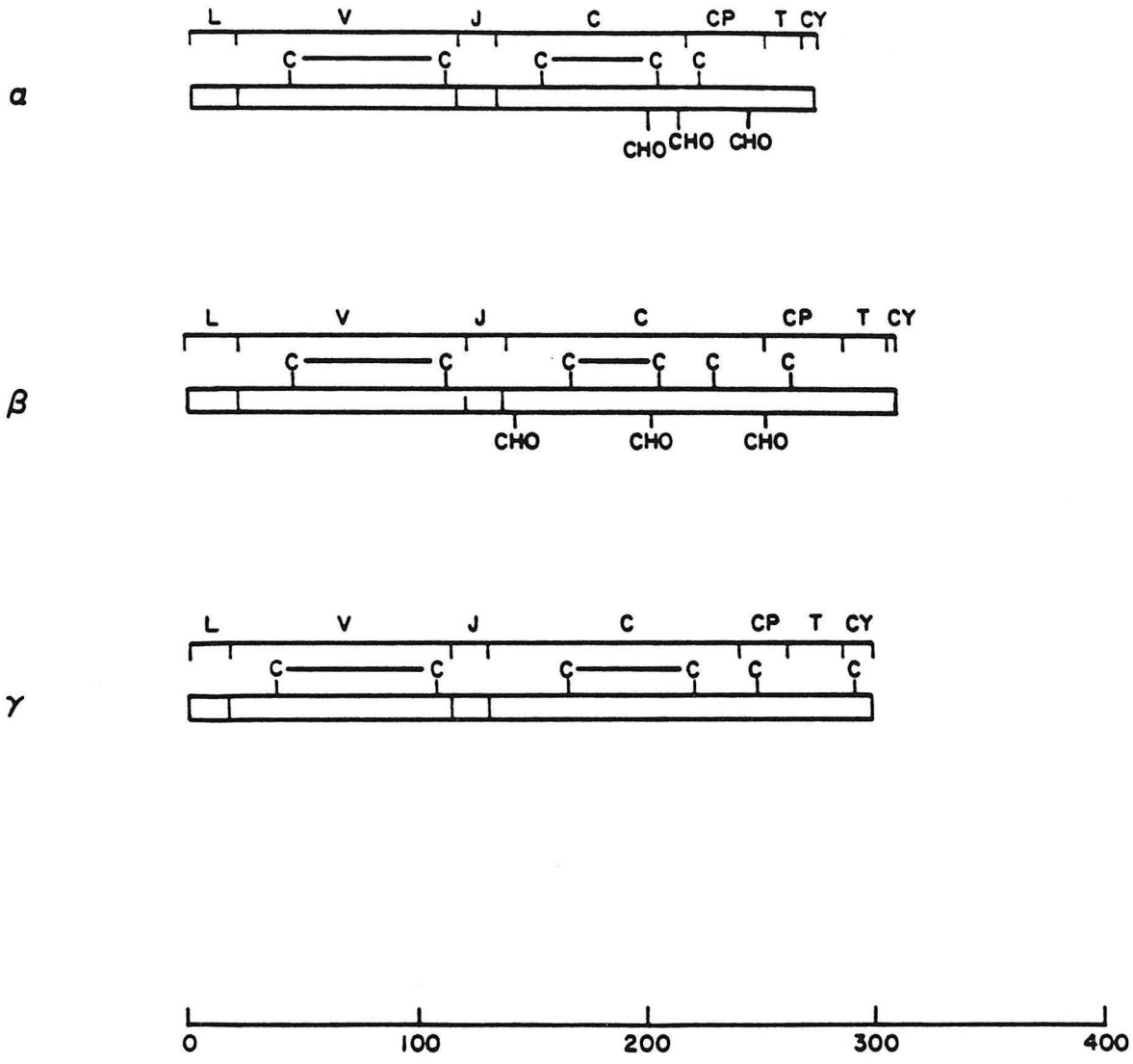
258. McDevitt, H. O., Deak, B. D., Shreffler, D. G., Klein, J., Stimpfling, J. H., Snell, G. D. 1972. Genetic control of the immune response. Mapping of the Ir-1 locus. *J. Exp. Med.* 135:1259-1278.

259. Schwartz, R. H. 1978. A clonal deletion model for Ir gene control of the immune response. *Scand. J. Immunol.* 7:3-10.

260. Shastri, N., Oki, A., Miller, A., Sercarz, E. 1985. Distinct recognition phenotypes exist for T cell clones specific for small peptide regions of proteins. Implications for the mechanism underlying major histocompatibility complex-restricted antigen recognition and clonal deletion models of immune response gene defects. *J. Exp. Med.* 162:332-345.

261. Davis, M. M. 1985. Molecular genetics of the T-cell receptor beta chain. *Ann. Rev. Immunol.* 3:537-560.

Figure 1. α, β and γ polypeptides. The murine α, β and γ chains are represented by open rectangles. The indicated structural features of the α, β and putative γ chain polypeptides were for the most part deduced from the nucleotide sequences of cDNA clones. Cysteines are represented by a C connected by a vertical line to the rectangle. The likely intrachain disulfide bridges are indicated by a horizontal line connecting two cysteine residues, and potential sites for N-linked glycosylation are represented by -CHO. Potential sites for N-linked glycosylation are also present in some V regions; these sites are not shown in the figure. Above the protein, the relative lengths of seven sections are indicated. L = leader, V = variable segment, J = joining segment, C = constant region, CP = connecting peptide, T = transmembrane region and CY = cytoplasmic region. The scale is given in amino acids.



Scale in amino acids

Figure 2. Genomic organization of the six known rearranging gene families of B and T cells. a. Murine immunoglobulin genes. b. Murine T-cell α, β and γ genes. Coding sequences are indicated by boxes for V gene segments and C genes or short vertical lines for D and J gene segments and for the exon encoding the leader of V gene segments; introns and flanking sequences are represented by a thin horizontal line. The separate exons and introns of the different C genes are not shown (see Figure 7). A double diagonal line indicates that all the flanking sequences are not depicted. The distance between two coding sequences is indicated in those cases where the intergene distance is known and all the flanking sequence is not shown. The nomenclature of Davis and colleagues has $C_{\beta 1} = C_T\beta$ and $C_{\beta 2} = C_T\beta'$ (261). The three murine C_γ genes have not been physically linked and the nomenclature for these genes is arbitrary. For the V gene segments with the exception of $V_{\beta 14}$ numbering is also arbitrary, in general the order of these gene segments is not known.

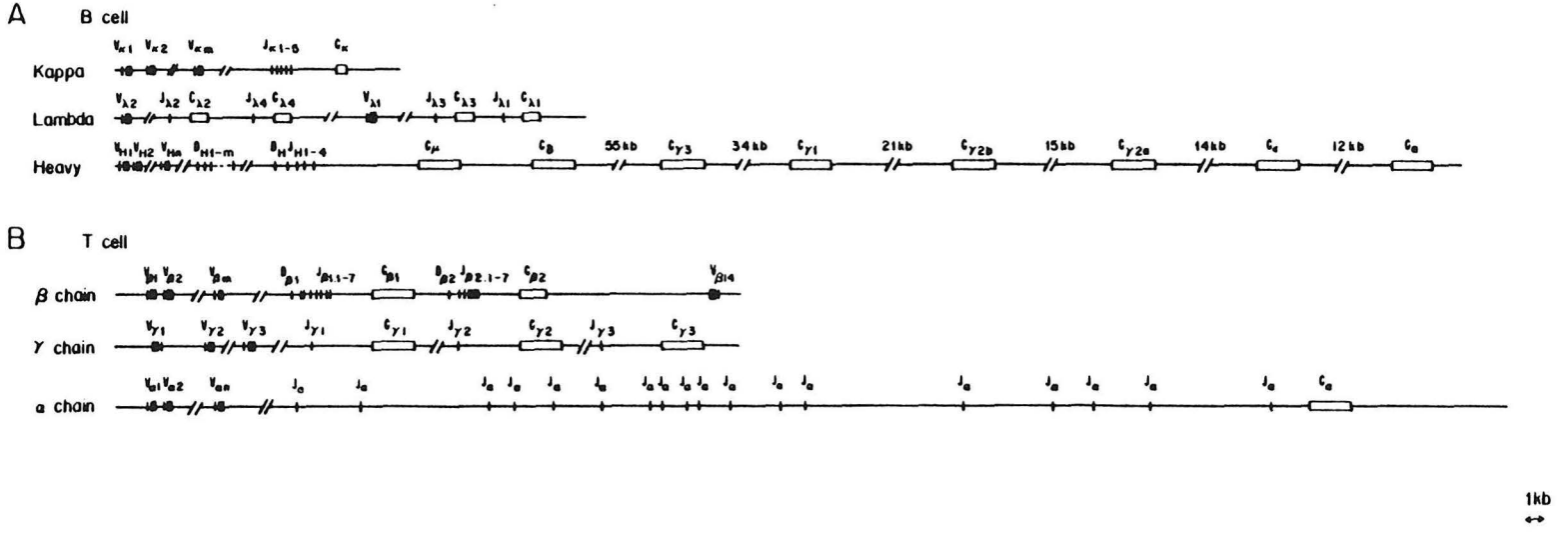
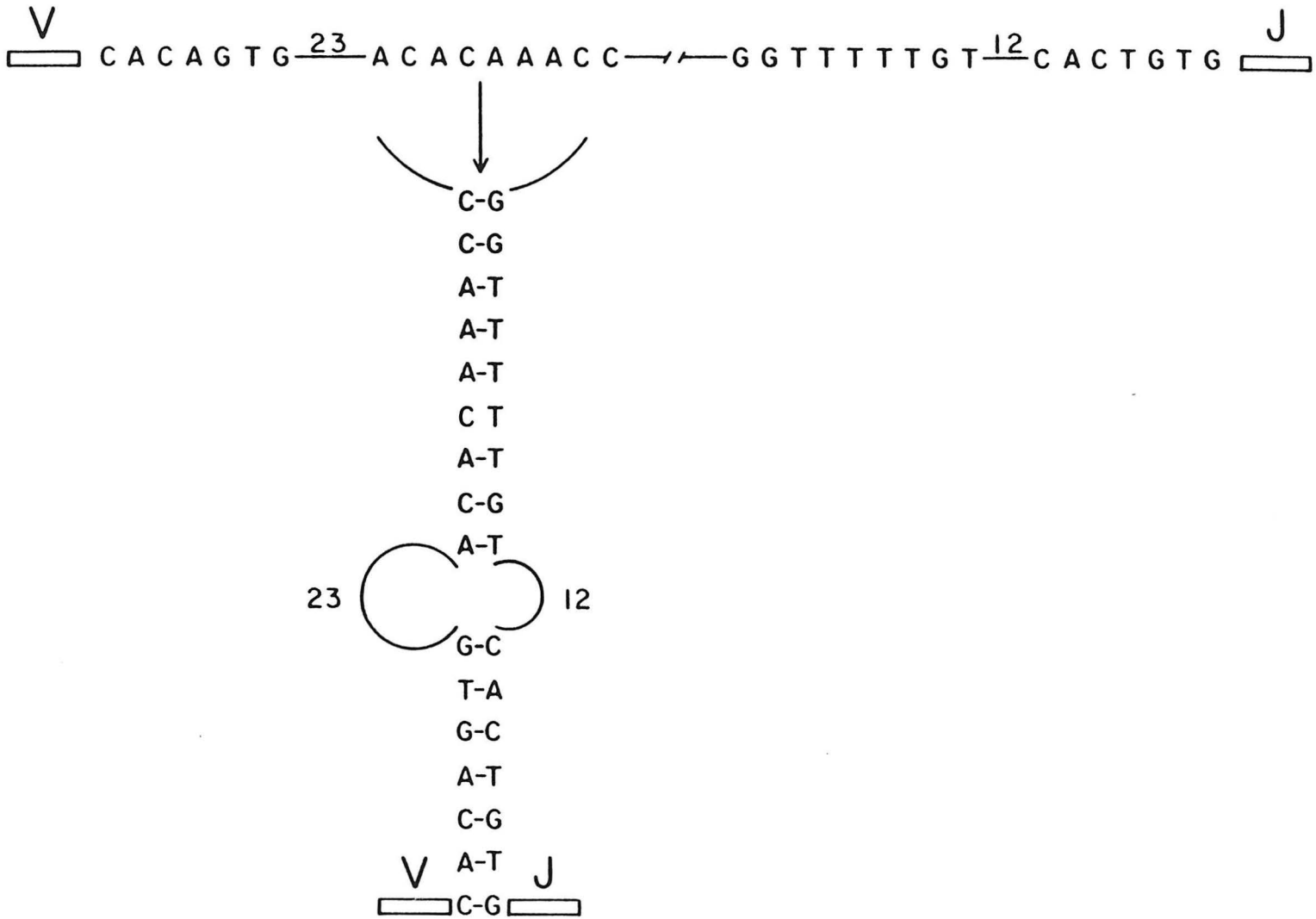

A
B cell
Kappa
Lambda
Heavy
55kb
34kb
21kb
15kb
14kb
12kb
B
T cell
β chain
γ chain
α chain
1kb

Figure 3. a. Rearrangement signals in a hypothetical V to J joining. All sequences are indicated by a thin line except for the conserved heptamer and nonamer, only the 'coding' strand of these sequences are shown. The 23 and 12 refer to the lengths of the nonconserved spacer sequence. The inverse complementarity and the ability of a single strand to form a base-paired stem-loop structure are illustrated. A base-paired structure of the type shown is a hypothetical intermediate in the recombination process. 3b. Rearrangement signals for DNA rearrangement in the six known rearranging gene families. The numbers 7 and 9 represent the conserved heptamer and A/T rich nonamer sequences, respectively. (1) denotes the one-turn or 12-base pair nonconserved spacer sequence, and (2) denotes the two-turn or 23 base pair spacer sequence. The uncertainty of the existence of D_{α} and D_{γ} gene segments is indicated by a question mark.


V
CACAGTG
23
ACACAAACC
GGTTTTTTGT
12
CACTGTG
J
C-G
C-G
A-T
A-T
A-T
C T
A-T
C-G
A-T
23
12
G-C
T-A
G-C
A-T
C-G
A-T
V
J
C-G

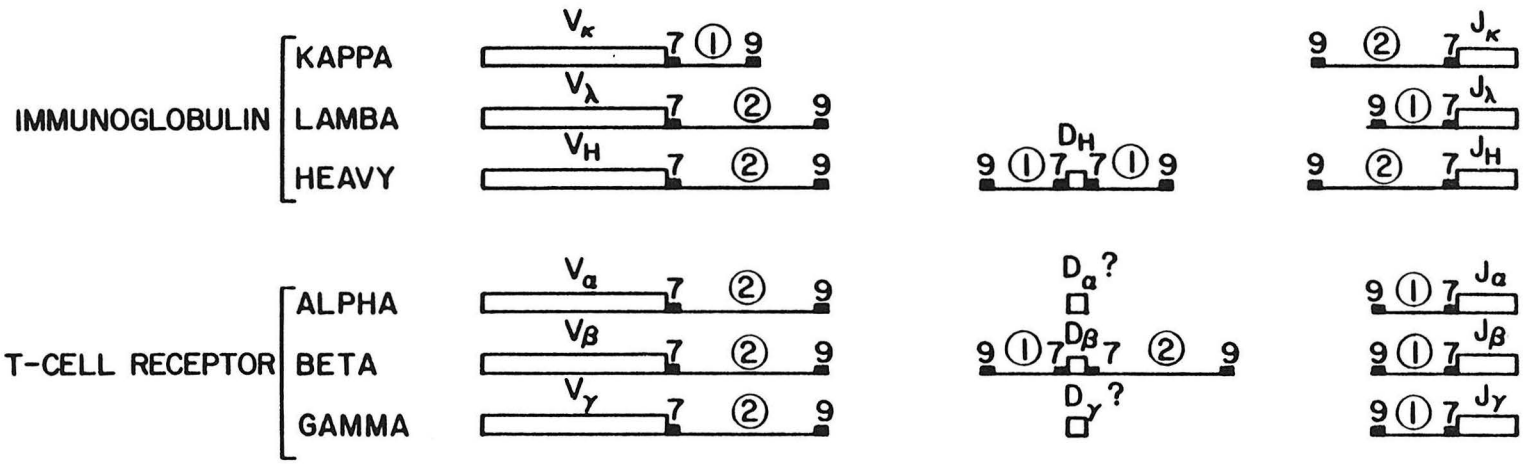

IMMUNOGLOBULIN
KAPPA
LAMBA
HEAVY
T-CELL RECEPTOR
ALPHA
BETA
GAMMA
Vκ
Vλ
VH
Vα
Vβ
Vγ
DH
Dα?
Dβ
Dγ?
Jκ
Jλ
JH
Jα
Jβ
Jγ
7
9
1
2

Figure 4. V_β segments. The protein sequences of 16 mouse V_β segments aligned for maximum similarity. The sequence data are derived from the references indicated (31, 32, 34, 86-89). The sequences of $V_{\beta 10}$, $V_{\beta 12}$ and $V_{\beta 13}$ are derived from cDNA clones that do not contain the complete V_β gene segments. A simplified but arbitrary nomenclature is used. These names are equivalent to the following designations from other papers. $V_{\beta 1}$ in Figure 4 = 86T1 (34) and $V_{\beta 11}$ (89); $V_{\beta 2}$ = E1 (87) and $V_{\beta 6}$ (89); $V_{\beta 3}$ = 2B4 (32) and $V_{\beta 13}$ (89); $V_{\beta 4} = V_{\beta 9}$ (89), $V_{\beta 5} = V_{\beta 8}$ (89), $V_{\beta 6}$ = LB2 (87) and $V_{\beta 1}$; $V_{\beta 7}$ = HDS11 (31) and $V_{\beta 14}$ (89); $V_{\beta 8.1}$ = C5 (87) and $V_{\beta 12}$ (89). $V_{\beta 8.2} = V_{\beta 4}$ (89); $V_{\beta 9} = V_{\beta 2}$ (89); $V_{\beta 10} = V_{\beta 3}$ (89); $V_{\beta 11} = V_{\beta 5}$ (89); $V_{\beta 12} = V_{\beta 7}$ (89); and $V_{\beta 13} = V_{\beta 10}$ (89). Members of the same subgroup are indicated by a fractional subgroup number, e.g., the $V_\beta 8.1$, $V_\beta 8.2$ and $V_\beta 8.3$ are members of a single $V_{\beta 8}$ subfamily. Conserved amino acids and amino acids important for V_H-V_L interaction are indicated. Conserved positions (>75% frequency of a single amino acid) are indicated by an asterisk if this amino acid is conserved only in the β gene family. If the same amino acid is conserved in other gene families as well, this is indicated by the letters in the open box. L = κ and λ; H = heavy; α = α chain. The arrow indicates invariant or semi-invariant amino acids that are thought to interact with one another to stabilize V_H-V_L interactions in immunoglobulins.

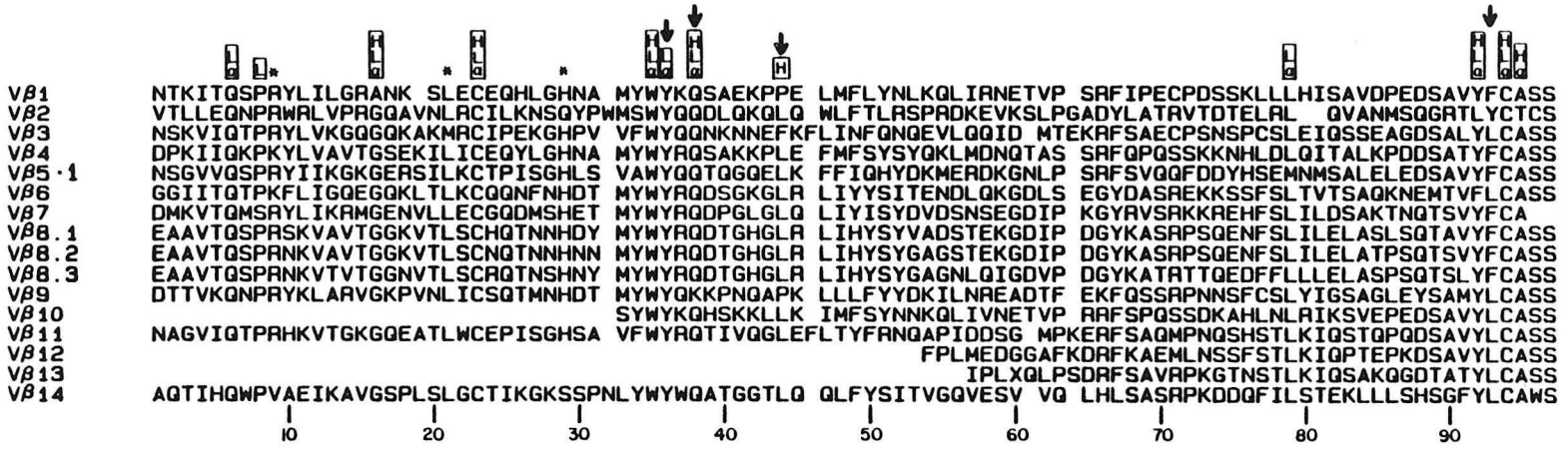

Vβ1 NTKITQSPRYLILGRANK SLECEQHLGHNA MYWYKQSAEKPPE LMFLYNLKQLIRNETVP SRFIPECPDSSKLLLHISAVDPEDSAVYFCASS
Vβ2 VTLLEQNPRWRLVPRGQAVNLRCILKNSQYPWMSWYQQDLQKQLQ WLFTLRSPRDKEVKSLPGADYLATRVTDTELRL QVANMSQGRTLYCTCS
Vβ3 NSKVIQTPRYLVKGQGQKAKMRCIPEKGHPV VFWYQQNKNNEFKFLINFQNQEVLQQID MTEKRFSAECPSNSPCSLEIQSSEAGDSALYLCASS
Vβ4 DPKIIQKPKYLVAVTGSEKILICEQYLGHNA MYWYRQSAKKPLE FMFSYSYQKLMDNQTAS SRFQPQSSKKNHLDLQITALKPDDSATYFCASS
Vβ5·1 NSGVVQSPRYIIKGKGERSILKCTPISGHLS VAWYQQTQGQELK FFIQHYDKMERDKGNLP SRFSVQQFDDYHSEMNMSALELEDSAVYFCASS
Vβ6 GGIITQTPKFLIGQEGQKLTLKCQQNFNHDT MYWYRQDSGKGLR LIYYSITENDLQKGDLS EGYDASREKKSSFSLTVTSAQKNEMTVFLCASS
Vβ7 DMKVTQMSRYLIKRMGENVLLECGQDMSHET MYWYRQDPGLGLQ LIYISYDVDSNSEGDIP KGYRVSRKKREHFSLILDSAKTNQTSVYFCA
Vβ8.1 EAAVTQSPRSKVAVTGGKVTLSCHQTNNHDY MYWYRQDTGHGLR LIHYSYVADSTEKGDIP DGYKASRPSQENFSLILELASLSQTAVYFCASS
Vβ8.2 EAAVTQSPRNKVAVTGGKVTLSCNQTNNHNN MYWYRQDTGHGLR LIHYSYGAGSTEKGDIP DGYKASRPSQENFSLILELATPSQTSVYFCASS
Vβ8.3 EAAVTQSPRNKVTVTGGNVTLSCRQTNSHNY MYWYRQDTGHGLR LIHYSYGAGNLQIGDVP DGYKATRTTQEDFFLLLELASPSQTSLYFCASS
Vβ9 DTTVKQNPRYKLARVGKPVNLICSQTMNHDT MYWYQKKPNQAPK LLLFYYDKILNREADTF EKFQSSRPNNSFCSLYIGSAGLEYSAMYLCASS
Vβ10 SYWYKQHSKKLLK IMFSYNNKQLIVNETVP RRFSPQSSDKAHLNLRIKSVEPEDSAVYLCASS
Vβ11 NAGVIQTPRHKVTGKGQEATLWCEPISGHSA VFWYRQTIVQGLEFLTYFRNQAPIDDSG MPKERFSAQMPNQSHSTLKIQSTQPQDSAVYLCASS
Vβ12 FPLMEDGGAFKDRFKAEMLNSSFSTLKIQPTEPKDSAVYLCASS
Vβ13 IPLXQLPSDRFSAVRPKGTNSTLKIQSAKQGDTATYLCASS
Vβ14 AQTIHQWPVAEIKAVGSPLSLGCTIKGKSSPNLYWYWQATGGTLQ QLFYSITVGQVESV VQ LHLSASRPKDDQFILSTEKLLLSHSGFYLCAWS
10
20
30
40
50
60
70
80
90

Figure 5. V_α segments. The protein sequences of 22 mouse V_α gene segments aligned for maximum similarity. The sequence data are derived from the references indicated (31, 33, 94, 95). The sequences are translated from the nucleotide sequences of cDNA clones, some of which were not full-length and therefore do not contain a complete V_α gene segment. The TA26 cDNA clone has an internal deletion and it is not known if this represents a cloning artifact. A simplified but arbitrary nomenclature is used as described for the V_β gene segment sequences. Individual V_α gene segments were not numbered, because in some cases it was not clear whether two closely related V_α gene segments were separate members of a subfamily, alleles that differ between mouse strains, or possibly somatic mutants. Conserved amino acid and amino acids important for V_H-V_L interactions are indicated as described in the legend to Figure 4.

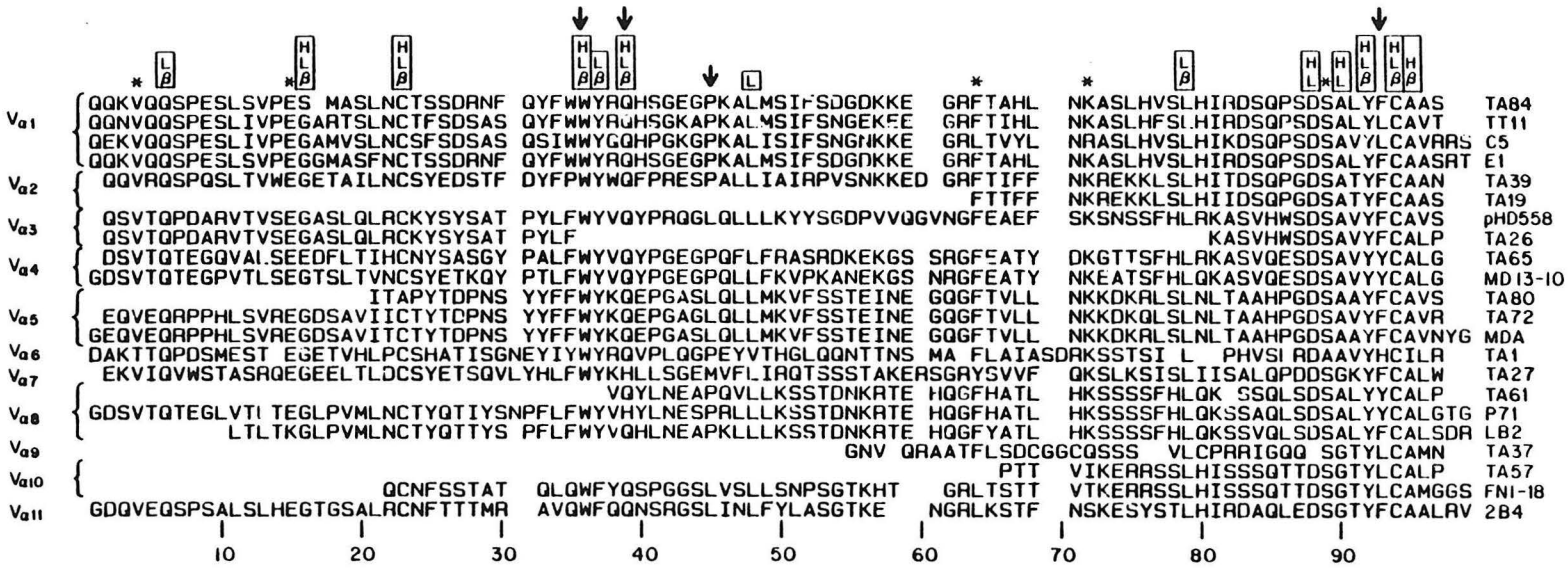

Vα1
Vα2
Vα3
Vα4
Vα5
Vα6
Vα7
Vα8
Vα9
Vα10
Vα11
QQKVQQSPESLSVPES MASLNCTSSDRNF QYFWWYRQHSGEGPKALMSIFSDGDKKE GRFTAHL NKASLHVSLHIRDSQPSDSALYFCAAS TA84
QQNVQQSPESLIVPEGARTSLNCTFSDSAS QYFWWYRQHSGKAPKALMSIFSNGEKEE GRFTIHL NKASLHFSLHIRDSQPSDSALYLCAVT TT11
QEKVQQSPESLIVPEGAMVSLNCSFSDSAS QSIWWYGQHPGKGPKALISIFSNGNKKE GRLTVYL NRASLHVSLHIKDSQPSDSAVYLCAVRAS C5
QQKVQQSPESLSVPEGGMASFNCTSSDRNF QYFWWYRQHSGEGPKALMSIFSDGDKKE GRFTAHL NKASLHVSLHIRDSQPSDSALYFCAASAT E1
QQVRQSPQSLTVWEGETAILNCSYEDSTF DYFPWYWQFPRESPALLIAIRPVSNKKED GRFTIFF NKREKKLSLHITDSQPGDSATYFCAAN TA39
FTTFF NKREKKLSLHIIDSQPGDSATYFCAAS TA19
QSVTQPDARVTVSEGASLQLRCKYSYSAT PYLFWYVQYPRQGLQLLLKYYSGDPVVQGVNGFEAEF SKSNSSFHLRKASVHWSDSAVYFCAVS pHD558
QSVTQPDARVTVSEGASLQLRCKYSYSAT PYLF KASVHWSDSAVYFCALP TA26
DSVTQTEGQVALSEEDFLTIHCNYSASGY PALFWYVQYPGEGPQFLFRASRDKEKGS SRGFEATY DKGTTSFHLRKASVQESDSAVYYCALG TA65
GDSVTQTEGPVTLSEGTSLTVNCSYETKQY PTLFWYVQYPGEGPQLLFKVPKANEKGS NRGFEATY NKEATSFHLQKASVQESDSAVYYCALG MD13-10
ITAPYTDPNS YYFFWYKQEPGASLQLLMKVFSSTEINE GQGFTVLL NKKDKRLSLNLTAAHPGDSAAYFCAVS TA80
EQVEQRPPHLSVREGDSAVIICTYTDPNS YYFFWYKQEPGAGLQLLMKVFSSTEINE GQGFTVLL NKKDKQLSLNLTAAHPGDSAVYFCAVR TA72
GEQVEQRPPHLSVREGDSAVITCTYTDPNS YYFFWYKQEPGASLQLLMKVFSSTEINE GQGFTVLL NKKDKRLSLNLTAAHPGDSAAYFCAVNYG MDA
DAKTTQPDSMEST EGETVHLPCSHATISGNEYIYWYRQVPLQGPEYVTHGLQQNTTNS MA FLAIASDRKSSTSI L PHVSI RDAAVYHCILR TA1
EKVIQVWSTASRQEGEELTLDCSYETSQVLYHLFWYKHLLSGEMVFLIRQTSSSTAKERSGRYSVVF QKSLKSISLIISALQPDDSGKYFCALW TA27
VQYLNEAPQVLLKSSTDNKRTE HQGFHATL HKSSSSFHLQK SSQLSDSALYYCALP TA61
GDSVTQTEGLVTLTEGLPVMLNCTYQTIYSNPFLFWYVHYLNESPRLLLKSSTDNKRTE HQGFHATL HKSSSSFHLQKSSAQLSDSALYYCALGTG P71
LTLTKGLPVMLNCTYQTTYS PFLFWYVQHLNEAPKLLLKSSTDNKRTE HQGFYATL HKSSSSFHLQKSSVQLSDSALYFCALSDR LB2
GNV QRAATFLSDCGGCQSSS VLCPRRIGQQ SGTYLCAMN TA37
PTT VIKERRSSLHISSSQTTDSGTYLCALP TA57
QCNFSSTAT QLQWFYQSPGGSLVSLLSNPSGTKHT GRLTSTT VTKERRSSLHISSSQTTDSGTYLCAMGGS FN1-18
GDQVEQSPSALSLHEGTGSALRCNFTTTMR AVQWFQQNSRGSLINLFYLASGTKE NGRLKSTF NSKESYSTLHIRDAQLEDSGTYFCAALRV 2B4
10
20
30
40
50
60
70
80
90

Figure 6. Comparison of murine J gene segments from six different gene families. In order from top to bottom, six J_α, six J_β, one J_γ, two J_H, two J_κ and two J_λ gene segments are shown. J_α gene segments are numbered according to the indicated references, the numbers are essentially clone numbers. Conserved amino acids found in most J gene segments are indicated by an asterisk. The position having conserved amino acids important for immunoglobulin V_H-V_L interaction is depicted by an arrow. Data are from the indicated references (71, 73, 75, 77, 78).

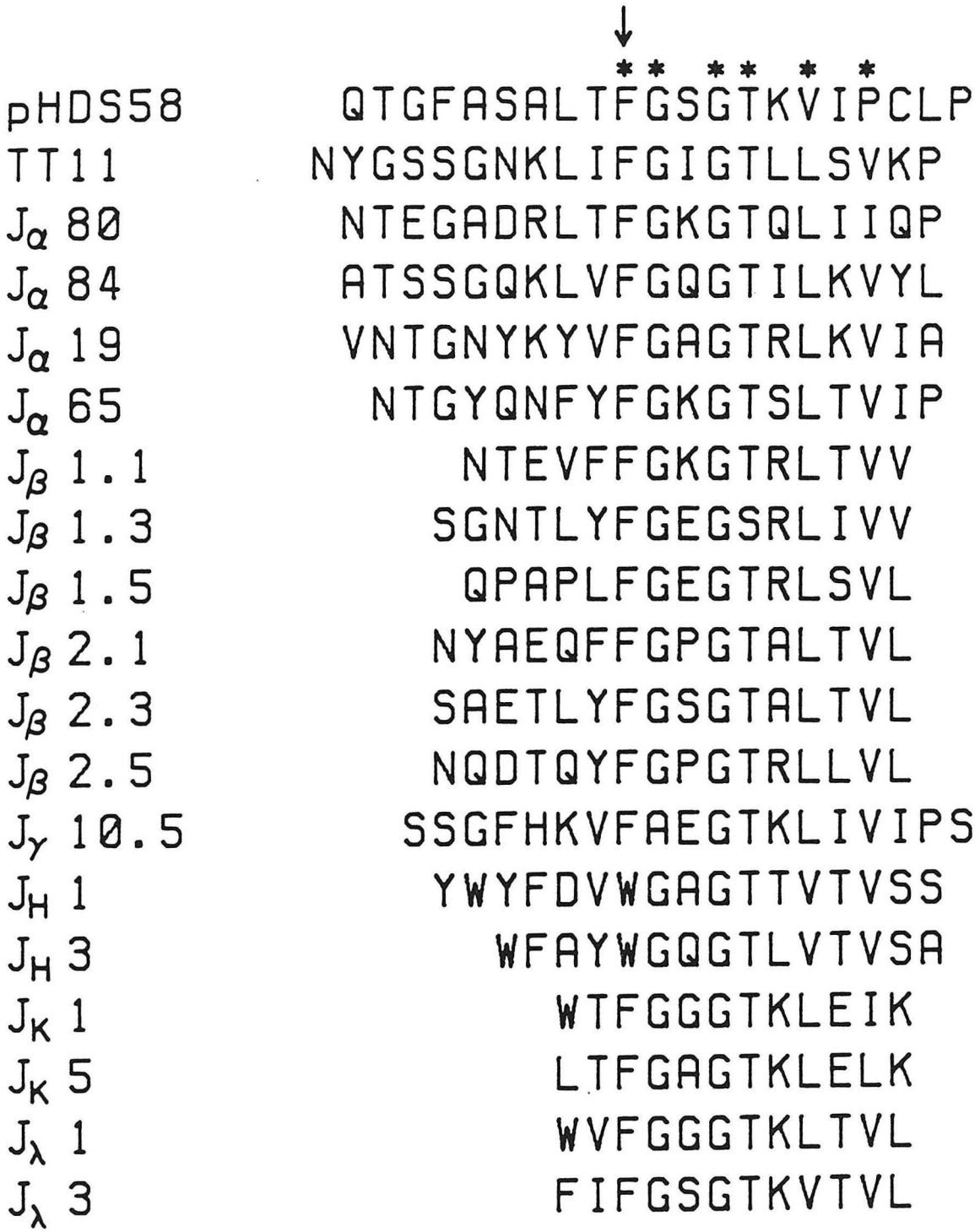

pHDS58 QTGFASALTFGSGTKVIPCLP
TT11 NYGSSGNKLIFGIGTLLSVKP
Jα 80 NTEGADRLTFGKGTQLIIQP
Jα 84 ATSSGQKLVFGQGTILKVYL
Jα 19 VNTGNYKYVFGAGTRLKVIA
Jα 65 NTGYQNFYFGKGTSLTVIP
Jβ 1.1 NTEVFFGKGTRLTVV
Jβ 1.3 SGNTLYFGEGSRLIVV
Jβ 1.5 QPAPLFGEGTRLSVL
Jβ 2.1 NYAEQFFGPGTALTVL
Jβ 2.3 SAETLYFGSGTALTVL
Jβ 2.5 NQDTQYFGPGTRLLVL
Jγ 10.5 SSGFHKVFAEGTKLIVIPS
JH 1 YWYFDVWGAGTTVTVSS
JH 3 WFAYWGQGTLVTVSA
JK 1 WTFGGGTKLEIK
JK 5 LTFGAGTKLELK
Jλ 1 WVFGGGTKLTVL
Jλ 3 FIFGSGTKVTVL

Figure 7. The genomic organization of a C_α, C_β and C_γ genes. Exons, numbered with Roman numerals, are depicted by boxes and introns by a horizontal line. Open boxes represent coding sequence and the hatched boxes represent 3' nontranslated sequences. The human C_α gene (76), the murine $C_{\beta 2}$ gene (77, 78) and a murine C_γ gene from clone J-C10.5 (73) are indicated. The scale in kilobases is given.

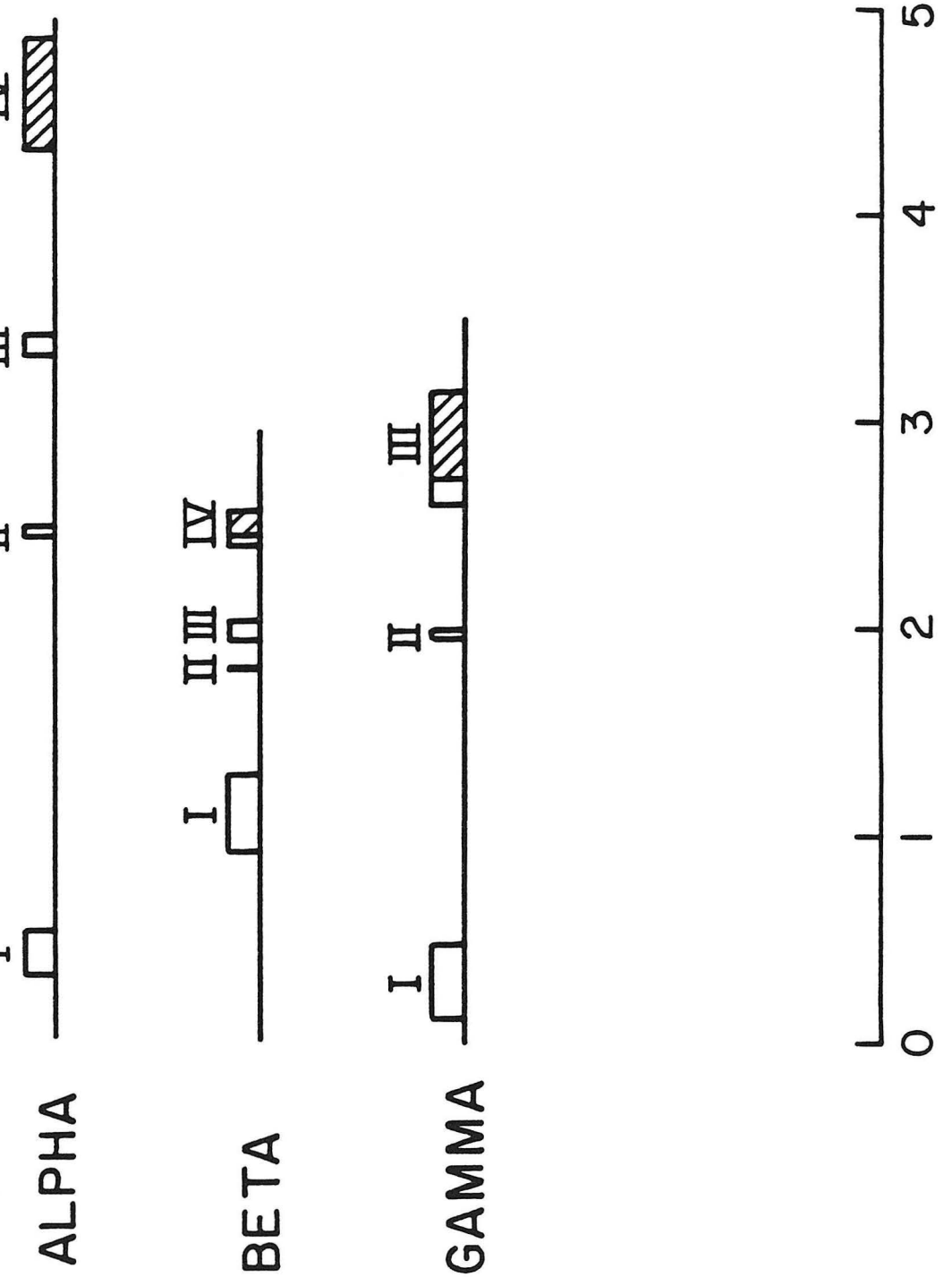

ALPHA
BETA
GAMMA
I
II
III
IV
0
1
2
3
4
5
Scale in kilobases

Figure 8. A schematic representation of the evolution of the immunoglobulin gene superfamily. The order of divergences is inferred from sequence homology and exon/intron structure (109). V and C denote a V- or C-like homology unit, respectively. The open circles for the MHC molecules do not exhibit sequence similarity with the immunoglobulin homology units, although they are of similar length. The Thy-1 homology unit does exhibit sequence similarity, although it is not easily classified as V or C and may have diverged prior to the V-C divergence. The horizontally paired homology units represent established domain structures, apart from those for the poly Ig receptor and the T8 molecule. The T4 gene and MRC-OX2 each have both a single V and a single C homology unit (not shown).

Immunoglobulin Gene Superfamily

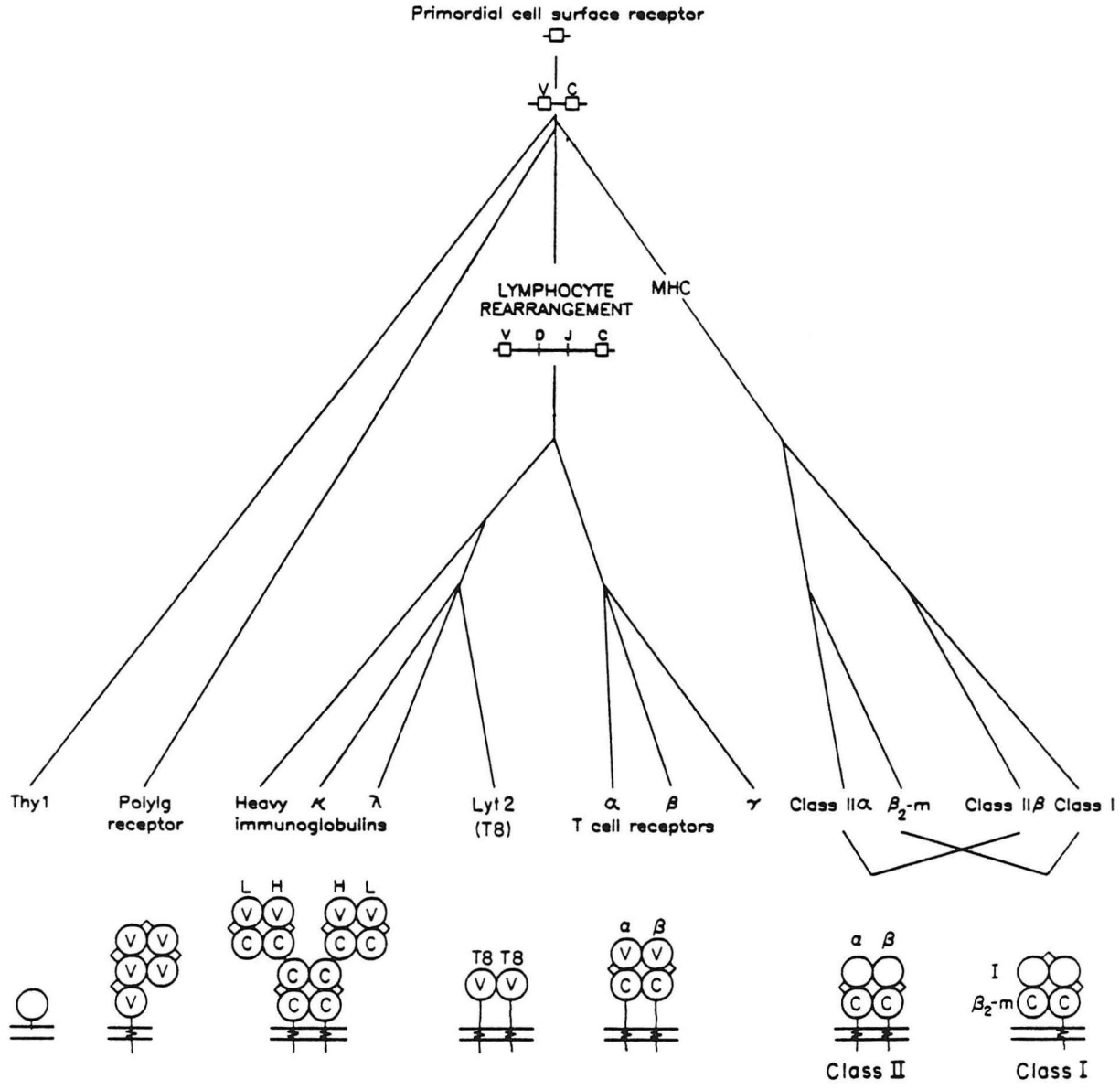
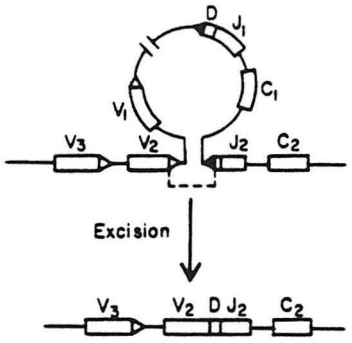
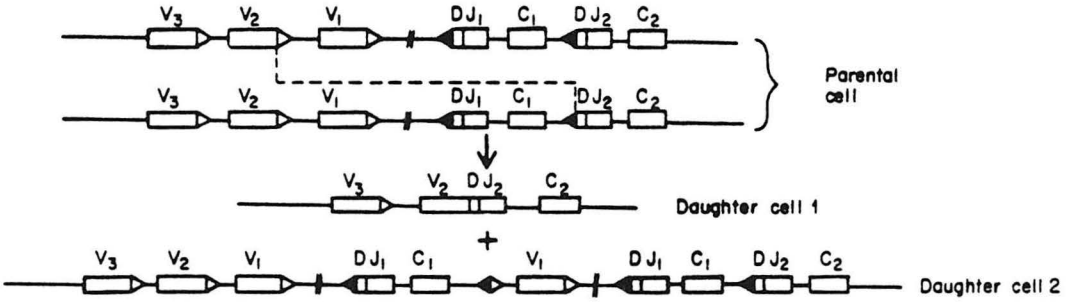


Figure 9. Three models for antigen receptor gene rearrangement. Rearrangements are shown schematically in a gene family with organizational features similar to the β gene family. The V, D and J gene segments and C genes, represented by boxes, are not drawn to scale. ▵ = one-turn recognition signals for DNA rearrangement. ▲ = two-turn recognition signals for DNA rearrangement. The double diagonal line indicates V gene segments that have not been physically linked to D or J gene segments. For clarity, only rearrangement of a hypothetical V2 gene segment to an already-joined D-J2 gene segment is illustrated, although the same models apply to D-J rearrangements. a. Excision or deletion. A stem and loop is formed by base-paired DNA joining recognition sequences; this stem-loop is then excised and the adjacent sequences, connected by a broken line, are joined. b. Unequal sister chromatid exchange. The right-hand brackets represent the connection of the sister chromatids at the centromere. The broken line connects the breakpoints in the two chromatids. The sequences duplicated on one of the two segregated chromosomes are underlined. The diamond on the horizontal line between C1 and V1 denotes a recombination breakpoint containing DNA rearrangement signals joined in a head-to-head fashion. c. Inversion. The horizontal arrows indicate the transcriptional orientation of the various gene segments and C genes. This figure is adapted from reference 144.

a. Excision



b. Sister chromatid exchange



c. Inversion

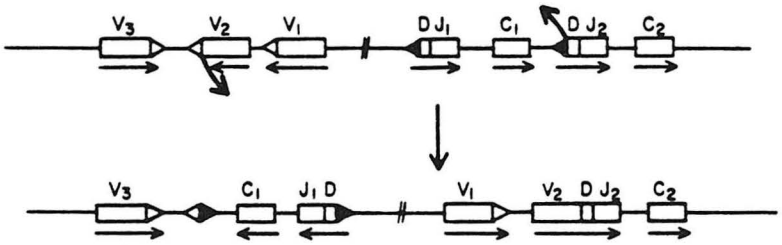


Figure 10. β gene rearrangement by two mechanisms. The figure depicts the likely sequence and mechanisms of β gene rearrangement in a T helper cell line (86). The V_β, D_β and J_β gene segments and C_β genes are indicated by open boxes, flanking sequences by thin horizontal lines. The separate exons of the two C_β genes are not shown. The drawing is not to scale. Δ and ∇ represent two-turn and one-turn DNA rearrangement signals, respectively. For simplicity, only some of the rearrangement signals adjacent to the gene segments are shown. ◀▶ represents the joined rearrangement signals that result from the inversion, and the horizontal arrows indicate the transcriptional orientation of some of the relevant coding sequences.

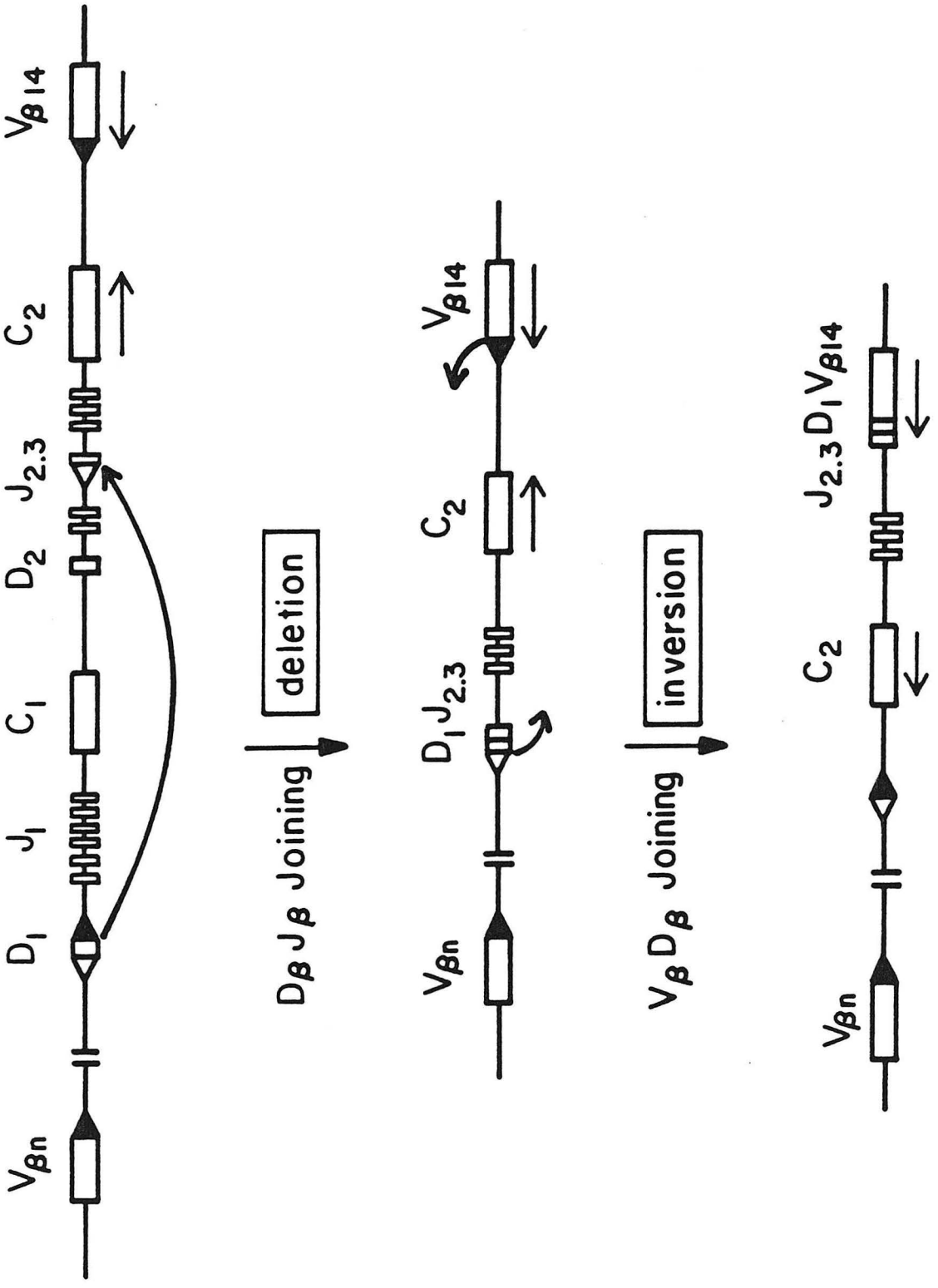

$V_{\beta 14}$
C_2
$J_{2.3}$
D_2
C_1
J_1
D_1
$V_{\beta n}$
$D_\beta J_\beta$ Joining
deletion
$V_{\beta 14}$
C_2
$D_1 J_{2.3}$
$V_{\beta n}$
$V_\beta D_\beta$ Joining
inversion
$J_{2.3} D_1 V_{\beta 14}$
C_2
$V_{\beta n}$

Figure 11. A schematic outline of four stages in the ontogeny of the rearrangement and expression of α, β and γ genes. A zero superscript, as in γ^0, represents a germline γ gene. A + superscript signifies productive rearrangement and a - superscript nonproductive rearrangement. The filled circle superscript represents a D-J rearrangement alone. The status of each homologue is illustrated; therefore γ^+/γ^- represents a cell with one productive and one nonproductive γ gene rearrangement. The status of the gene rearranging at a particular stage is presented. Class I, T_C = MHC class I-restricted cytotoxic T cell. Class II, T_H = MHC class II-restricted helper T cell.

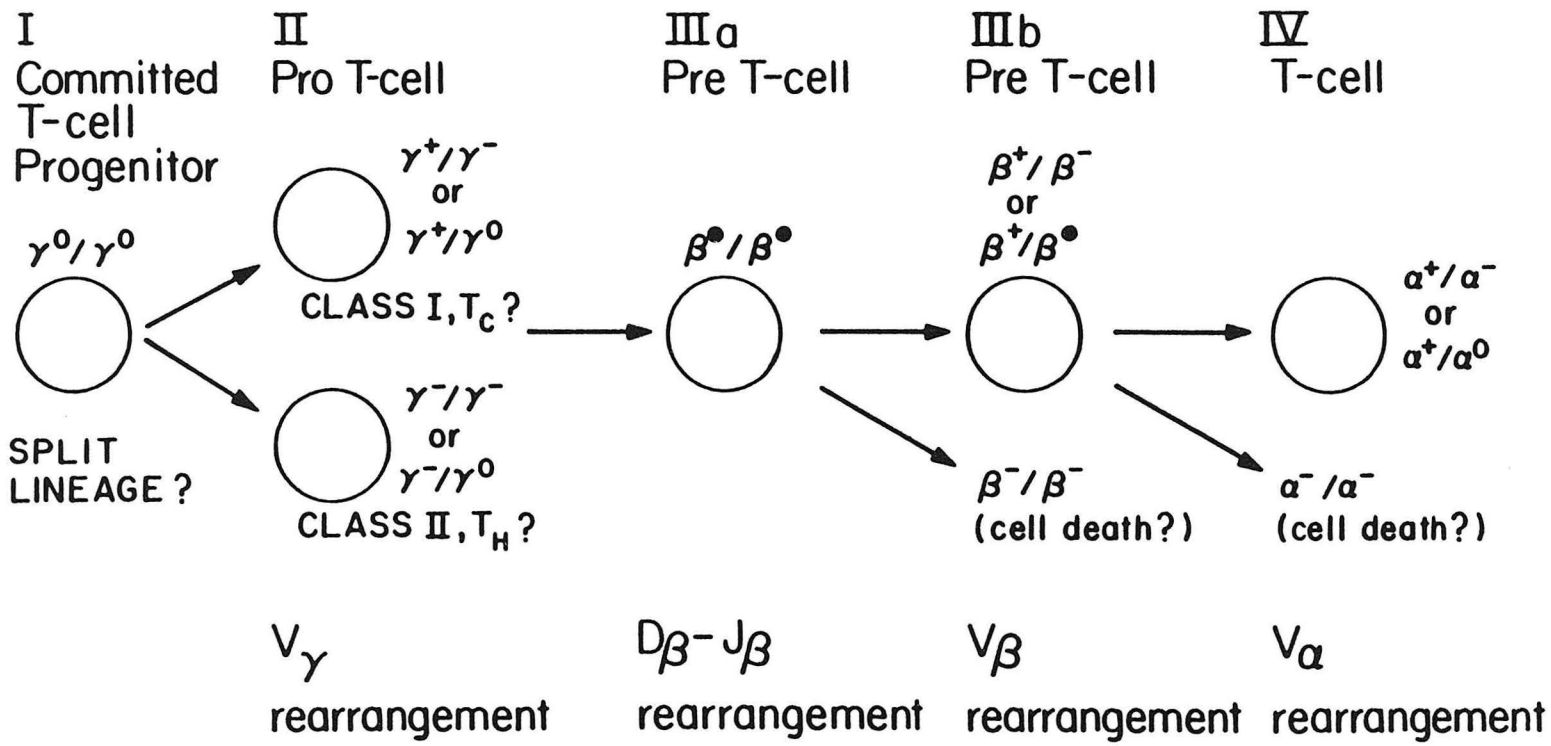

I
Committed
T-cell
Progenitor
II
Pro T-cell
IIIa
Pre T-cell
IIIb
Pre T-cell
IV
T-cell
γ0/γ0
SPLIT
LINEAGE?
γ+/γ−
or
γ+/γ0
CLASS I, Tc?
γ−/γ−
or
γ−/γ0
CLASS II, TH?
β●/β●
β+/β−
or
β+/β●
β−/β−
(cell death?)
α+/α−
or
α+/α0
α−/α−
(cell death?)
Vγ
rearrangement
Dβ−Jβ
rearrangement
Vβ
rearrangement
Vα
rearrangement

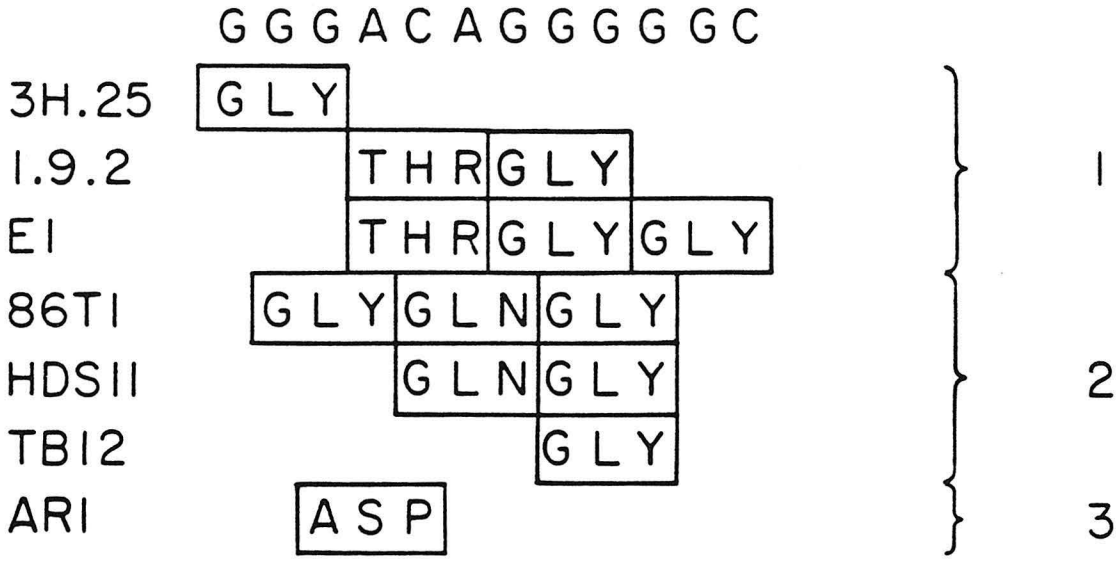
Figure 12. Junctional and N-region diversity in functional V_β genes. Two rearrangements that include the $V_\beta 3$ gene segment but different D_β and J_β gene segments are illustrated; one from a lysozyme/I-A^b specific T_H clone (3H.25), and one from a cytochrome c/I-E^k specific T_H hybridoma (2B4). The rearranged and germline sequences are from the indicated references (71, 77, 78, 81, 82, 148). For each rearranged gene, the germline gene segments that comprise it are listed above, and the rearranged sequence below. For the rearranged sequence, nucleotides introduced by N-region diversification that are not in any of the germline gene segments are indicated, while nucleotides encoded in the germline are represented by a dot. A blank beneath a nucleotide encoded in the germline indicates those sequences deleted by the rearrangement event (junctional diversity).

	$V_{\beta 3}$	$D_{\beta 1}$	$J_{\beta 1\text{-}2}$
GERMLINE	CAGCAGTCTGTC	GGGACAGGGGGCC	CAAACTCC
3H.25	.....	TC.....	....

	$V_{\beta 3}$	$D_{\beta 2}$	$J_{\beta 2\text{-}5}$
GERMLINE	CAGCAGTCTGTC	GGGACTGGGGGGGCC	AACCAAGAC
2B4	.....	A.....A	.....

Figure 13. Ds in three. The D_β gene segments can be read in three reading frames. The nucleotide sequences of the two known D_β gene segments are shown along with the part of the D region amino acid sequence that could be encoded by germline D gene segments, as opposed to nucleotides due to N-region diversification, from 15 V_β genes. Sequences are from the indicated references (31, 32, 34, 87, 88, 148). The reading frame for each set of translated sequences is shown on the right. Those sequences in the same reading frame are enclosed by a bracket on the right, and the reading frame is indicated by a number.

Dβ1


G G G A C A G G G G G C
3H.25
G L Y
1.9.2
T H R G L Y
E1
T H R G L Y G L Y
86T1
G L Y G L N G L Y
HDS11
G L N G L Y
TB12
G L Y
AR1
A S P
1
2
3

Dβ2

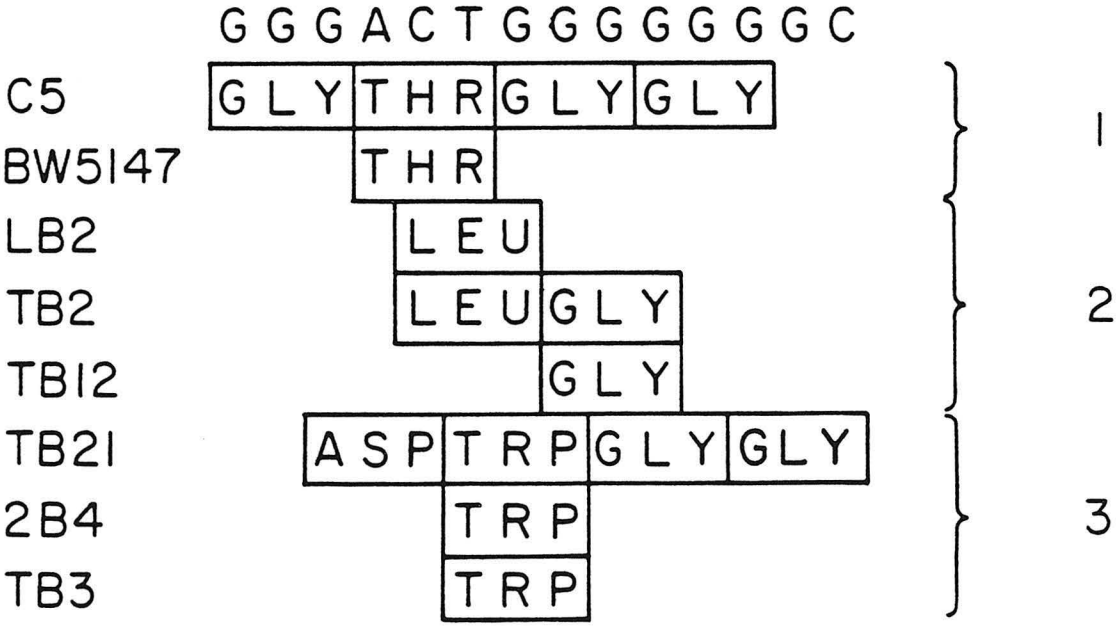

G G G A C T G G G G G G C
C5
G L Y T H R G L Y G L Y
BW5147
T H R
LB2
L E U
TB2
L E U G L Y
TB12
G L Y
TB21
A S P T R P G L Y G L Y
2B4
T R P
TB3
T R P
1
2
3

Table 1

Chromosomal Locations of the α, β and γ Genes

	Mouse	Human
α	14C-D	14q11-12
β	6B	7q32-35
γ	13A2-3	7p15

Details concerning the experimental methods and references are given in the text.

Table 2

Gene Rearrangement and Expression in Developing T Lymphocytes

	DNA Rearrangements				RNA Transcripts		
	γ		β		C_γ	C_β	C_α
	9.0 kb	16.0 kb	D-J	V-D-J		(1.3 kb)	(1.7 kb)
Fetal liver d12-19	+/- (d13 only)	+/- (d17-19 only)	+/-(?) (d13 only)	-	-	-	-
Fetal thymus							
d14	+	+/-	+/-	-	+	-	-
d15	+	+/-	+	+/-	+	+/-	-
d16	-	+	+	+	+	+	+/-
d17	-	+	+	+	+	+	+
Neonatal thymus	-	+	+	+	+	+	+
Adult thymus							
Lyt-2$^-$, L3T4$^-$	n.d.	n.d.	+	+	+	+	+/-
Lyt-2$^-$, L3T4, gp-1$^+$	n.d.	n.d.	-	-	n.d.	n.d.	n.d.

The genomic arrangement and expression of the α, β and γ genes in various developing populations is indicated. + = rearrangement or transcript is present. For detectable transcripts, different levels of abundance are not distinguished. +/- = rearrangement is present but in less than 10% of the chromosomes or the transcript is barely detectable. - = rearrangement or transcripts not detectable. n.d. = not done. The 16.0 kb γ gene rearrangement found in EcoRI cut DNA results from V_γ-J_γ joining and could be productive; the sequences responsible for the 9.0 kb γ gene rearrangement found in EcoRI digested RNA have not been characterized. The β gene rearrangement indicated by (?) is probably nonproductive, but it is not a D_β-J_β rearrangement. The presence or absence of the complete 1.3 kb C_β and 1.7 kb C_α transcripts are listed in the table; the shorter C_β and C_α transcripts are not included. The table is compiled from the references cited in the text. Information on other adult thymus subpopulations can be found in the indicated references (12, 173, 192, 193).

Table 3

Potential Diversity in Murine T-Cell Receptor and Immunoglobulin Variable Regions

	Immunoglobulins		T-Cell Receptor	
	H	L	β	α
Variable gene segment	250	250	30	100
Diversity gene segment	10	-	2	-
D in three reading frames	-	-	3	-
Joining gene segment	4	4	12	50
Combinatorial joining	$250\times10\times4 = 1\times10^4$	$250\times4 = 1\times10^3$	$(30\times1\times3\times12) + 30\times1\times3\times6 = 1620$*	$100\times50 = 5000$
Combinatorial association	$1\times10^4\times1\times10^3 = 1\times10^7$		$1620\times5000 = 8.1\times10^6$	

Table 3 Legend

* The calculation assumes the $D_{\beta1}$ rearranges to any of the 12 functional J_β gene segments, but that $D_{\beta2}$ rearranges only to the six functional $J_{\beta2}$ gene segments.

Table 4

Conserved Amino Acids Important for Immunoglobulin Variable Region Structure

Number	Amino acid	Position in immunoglobulins	Presence in T-cell genes α	β	γ	Intrachain folding (H bond)	Interchain (V_H-V_L interaction)
1*	CYS	23L, 22H	+	+	+	+	
2*	TRP	35L, 36H	+	+	+	+, with 12-13 positions C terminal	
3	TYR/VAL	TYR = 36L VAL = 37H	+ (TYR)	+ (TYR)	+ (TYR)	+, with TYR(PHE) 87L or 91H	+, with TRP 103H +, with PHE 98L
4*	GLN	38L, 39H	+	+	+	+, with amino acid 85L or 89H	+, with GLN 38L or 39H
5	PRO/LEU	PRO = 44L LEU = 45H	-[1]	+ (LEU)	-[1]	-	+, with TRP 103H +, with PHE 98L and TYR 87L
6	ASP	82L, 86H	+	-[2]	+	+, hydrogen bonds to amino acid 79L or 83H	+, with TRP 103H
7	ALA	84L, 88H	+	-[3]	+	+, with VAL 109H or VAL, LEU 104L	
8*	TYR	86L, 90H	+	+	+	+, with THR 102L or 107H	
9	TYR(PHE)[4]	87L, 91H	-[4]	-[4]	+	+, with TYR 36L or VAL 37H	+, with PRO 44L or LEU 45H
10*	CYS	88L, 92H	+	+	+	+, with glycine 99L or 104H	
11	PHE/TRP	PHE = 98L TRP = 103H	+ (PHE)	+ PHE)	+ PHE)		+, with TYR36L +, with VAL37H
12*	GLY	99L, 104H	+	+	-[5]	+, with CYS 88L or 92H	
13*	THR	102L, 107H	+	+	+	+, with TYR 86L or 90H	
14	VAL, LEU	104L, 109H	+	+	+	+, with ALA 84L or 88H	

Table 4 Legend

Fourteen relatively conserved amino acids that may be important for intrachain folding and/or V_H-V_L interaction of variable region homology units. Immunoglobulin sequence data and numbering of the positions is according to Kabat et al. (229) The first 10 positions are encoded by the V gene segment and the last four by the J gene segment. All conserved amino acids are present in >75% of both light and heavy chain sequences, except for positions 36L or 37H (#3), 44L or 45H (#5) and 98L or 103H (#11), where V_L and V_H each have a different conserved amino acid, and at 104L, where light chains have either VAL or LEU. The asterisk marks seven positions where the indicated amino acid is present in >95% of both V_H and V_L regions. Presence of the conserved amino acid in the V_γ gene from pHD54, and in >75% of the V_α, J_α, V_β and J_β gene segments is indicated by a + under the appropriate column. If a conserved amino acid is involved in an intrachain hydrogen bond in both the light and heavy chains of the myeloma protein NEW, this is indicated by a + under intrachain folding. If this H bond involves a second conserved amino acid on both chains, this is also indicated. Conserved amino acids that interact with other conserved amino acids to potentially stabilize V_H-V_L interactions are also indicated by a + under the appropriate column along with the amino acid involved in the interaction. Two amino acids are considered to be in contact if X-ray diffraction studies indicate that at least one of their atoms are within 1.2 van der Waals radii of one another. 1. 45α. Eighteen mouse alpha chains have 11 prolines, six leucines and one methionine (Figure 5); the V_γ gene has a phenylalanine. 2. 88β. Sixteen mouse β chains have eight aspartic acid, five glutamine and three other amino acids at this position (Figure 4). 3. 90β. Sixteen mouse V_β genes have 10 alanines, acids, three serines, and three other amino acids at this position (Figure 4). 4. Light chains have 70% tyrosine, 29% phenylalanine, 1% other and heavy chains have 80% tyrosine, 19% phenylalanine and 1% other amino acids. 92α and 93β. Twenty-two V_α genes have 12 phenylalanines, five leucines, four tyrosines and one histidine (Figure 5); 16 V_β genes have eight leucines, seven phenylalanines and one tyrosine. 5. Alanine is present in V_γ.

Table 5

V_β Genes Expressed by Functional Murine T Cells

T cell	Strain of origin	Specificity		V_β gene segments			Reference
		Antigen	MHC molecule	V_β	D_β	J_β	
1.9.2	B10.A(5R)	(Am-DAsp)	I-$E^{k/b}$	1	1	1.1	88
A20.2.15	(BALB/c x B6)F_1	Beef insulin	I-A^d	1	1		+
E1	BALB/c	trinitrophenyl	I-A^d	2	1	2.2	87
AR1	C57BL	alloreactive	H-2^{d*}	2	1	2.5	87
AO1T.8	B10.A	lysozyme (aa 74-86)	I-A^k	2	1	1.3	++
A10	B10.A	ovalbumin	I-A^k	2	1	1.2	86
2B4	B10.A	cytochrome	I-E^k	3	2	2.5	71
3H.25	C57BL/6	lysozyme	I-A^b	3	1	1.2	148
V11.5	B10A(5R)	cytochrome c	I-$E^{k/b}$	3	1	1.2	Δ
LB2	C67BL/6	chicken RBC	I-A^b	6	2	2.3	87
AOIC.25.1	B10.A	lysozyme (aa 85-96)	I-E^k	6	1or2	2.2	++
3F9	BALB/c	alloreactive	D^b	6	1or2	1.1	159
2C	BALB.B	alloreactive	D^d	7	1	2.6	31
C5	C57BL/6	DNP-ovalbumin	I-A^b	8.1	2	2.5	87
AOIT.13.1	B10.A	lysozyme (aa 85-96)	I-E^k	8.3	1	1.1	++
V15.4	B10.A(5R)	cytochrome c	I-$E^{k/b}$	8.3	1	1.4	Δ
AOIC.9.4	B10.A	lysozyme (aa 85-96)	I-E^k	10	1	1.2	++
AOIC.19.3	B10.A	lysozyme (aa 85-96)	I-E^k	10	1or2	2.2	++
J6-19	B10.A	ovalbumin	I-A^k	14	1	2.3	86

* Specific for an undefined MHC class I molecule.

+ T. Wegmann, personal communication.

++ J. Kobori and N. Shastri, unpublished data.

Δ J. Goverman, unpublished data.

APPENDIX I

THREE T-CELL HYBRIDOMAS DO NOT CONTAIN DETECTABLE HEAVY CHAIN VARIABLE GENE TRANSCRIPTS

Published in the *Journal of Experimental Medicine*

THREE T CELL HYBRIDOMAS DO NOT CONTAIN DETECTABLE HEAVY CHAIN VARIABLE GENE TRANSCRIPTS*

By MITCHELL KRONENBERG, ELLEN KRAIG, GERALD SIU, JUDITH A. KAPP,‡ JOHN KAPPLER,§ PHILIPPA MARRACK,§ CARL W. PIERCE,‡ AND LEROY HOOD

From the Division of Biology, California Institute of Technology, Pasadena, California 91125

There is considerable controversy as to whether or not the genes encoding the T cell antigen receptor are homologous to immunoglobulin gene segments. Results from a number of different experiments have been interpreted as evidence that T lymphocytes utilize heavy chain variable (V_H)[1] regions to bind specifically to antigen. For example, some antiidiotypic and anti-V_H framework sera interfere with T cell function and/or bind to antigen-specific factors secreted by T cells (1–3). In several cases, the gene encoding the cross-reactive determinant expressed by the T cells is linked to the immunoglobulin heavy chain gene cluster (4–9). Also, the genes encoding a series of T cell alloantigens have been mapped to chromosome 12, between the C_H gene locus, Igh-1, and the prealbumin gene (9, 10). Recently, these alloantigens have been detected on antigen-binding factors secreted by T cells (11–13). It has been proposed that the antigenic determinants encoded by C_H-linked genes are T cell isotypes that may be expressed in conjunction with V_H gene segments (10, 14). Furthermore, some T lymphocytes contain rearranged J_H gene segments or a C_μ transcript (15–23). This may indicate that the mechanisms controlling V_H-D-J_H joining and immunoglobulin transcription also operate upon homologous sequences in the synthesis of T cell antigen-binding receptors.

Experiments that report the expression of V_H serologic determinants by T lymphocytes have provided the most extensive and convincing data in support of V_H gene transcription by T cells. However, the serologic data are indirect, and there are three possible ways to interpret them. First, T and B cells

* Supported by National Institutes of Health grants AI 17565, AI 18959, AI 15353, AI 18785, and Public Health Service (PHS) grant no. IF 32CA06693 (to E. Kraig), awarded by the National Cancer Institute, DHHS.

‡ Depts. of Pathology and Laboratory Medicine, The Jewish Hospital of St. Louis and Depts. of Pathology and of Microbiology and Immunology, Washington University School of Medicine, St. Louis, Missouri 63110.

§ Dept. of Medicine, National Jewish Hospital and Research Center, Denver, Colorado 80206.

[1] *Abbreviations used in this paper*: cDNA, DNA complementary to mRNA; C_H, heavy chain constant region; C_μIgM constant region; D, diversity gene segment; GAT, L-glutamic acid^{60}-L-alanine^{30}-L-tyrosine^{10}; GT, L-glutamic acid^{50}-L-tyrosine^{50}; HGG, human gamma globulin; J_H, heavy chain joining gene segment; kb, kilobase; KLH, *keyhole limpet* hemocyanin; MHC, major histocompatibility complex; NP, 4-hydroxy-3-nitrophenyl acetyl; PC, phosphorylcholine; SET, 0.75 M NaCl/0.15 M Tris, pH 8.0/5 mM ethylenediaminetetraacetic acid; V_H, heavy chain variable; V_κ, kappa light chain variable; V_λ, lambda light chain variable.

 J. Exp. Med. © The Rockefeller University Press · 0022-1007/83/07/0210/18 $1.00
Volume 158 July 1983 210–227

responding to the same antigen may express highly similar or identical V_H gene segments. This is unlikely since, in several experimental systems the receptor synthesized by T cells responding to an antigen does not share all the serologic determinants present on the immunoglobulin synthesized by B cells responding to the same antigen (1, 24). In addition, we and others have demonstrated that idiotype-positive T and B lymphocytes that respond to the same antigen do not transcribe highly similar V_H gene segments (25, 26). Second, it is possible that T cells use the repertoire of V_H genes differently than B cells do. This might occur because T lymphocytes do not express light chain genes (3, 4) or because T cells recognize antigen in the context of syngeneic MHC gene products. If this were true, then T cells responding to an antigen may transcribe V_H gene segments that have limited structural similarity to those transcribed in B cells responding to the same antigen, although these different V_H gene products could share some idiotopes. Finally, it is possible that the V_H cross-reactive determinants present on T cells and T cell factors are not the products of V_H genes.

In this paper, we report our attempts to determine whether any V_H gene segments are expressed in T lymphocytes. cDNA libraries were constructed from a suppressor T cell hybridoma specific for the synthetic polypeptide GAT, and from two helper T cell hybridomas, one specific for HGG and the second responding to KLH. The cDNA libraries were hybridized with two sets of probes; each set capable of detecting a wide range of V_H gene segments. In constructing the probes, no assumptions were made concerning the degree of homology between the B cell heavy chain variable regions binding GAT, HGG, or KLH and V_H gene transcripts that might be present in the T cell hybridomas. One set of probes was a synthetic oligonucleotide complementary to a conserved sequence found at the 3′ end of many mouse V_H gene segments and a single-stranded cDNA synthesized primarily from the heavy chain variable genes present in spleen RNA. The second set of probes was two cloned V_H gene segments, one from the V_HII gene subgroup and one from the V_HIII gene subgroup. The cDNA libraries were sufficiently large so that the chance of detecting a sequence found in the nonabundant messenger RNA class (10–20 copies per cell) was excellent. Since no V_H-containing cDNA colonies were found, we conclude that V_H gene segments are not likely to encode the T cell antigen-binding receptor.

Methods

RNA Preparation. T cell hybridomas were grown in liquid culture and harvested. The cell pellets were lysed in guanidinium thiocyanate, and the RNA was prepared by centrifugation through a cushion of cesium chloride (27). The percent yield and the amount of RNA per cell were estimated using a recovery marker as previously described (20). RNA was similarly prepared from spleens of 6-mo old BALB/c Cum mice. Poly(A)$^+$ RNA was purified by two cycles of oligo(dT)-cellulose chromatography (28).

cDNA Synthesis. Double-stranded cDNA was synthesized as described (29, 30). First strand synthesis was initiated by random priming using sheared calf thymus DNA (31). The double-stranded cDNA was fractionated by gel filtration and the material ranging in size from 400 to 1,500 base pairs was pooled. The average length of the cDNA was ~800 base pairs. The cDNA was cloned into the Pst I site of the tetracycline-resistant plasmid pBR322 by annealing dC-tailed cDNA to dG-tailed vector (32). Bacterial strain MC1061 (33) was transformed with cDNA and the transformants were selected with tetracycline

(34). We obtained ~10^6 colonies per μg of cDNA. Transformation with vector alone (dG-tailed pBR322) yielded a 2% background.

Synthetic Oligonucleotides. Two undecamers were synthesized separately by Dr. S. Horvath (California Institute of Technology) by the phosphite coupling method (35, 36). The sequences were verified by the method of Maxam and Gilbert (37). The oligonucleotide probes were labeled with 5′-[γ-^{32}P]dATP to a specific activity of >2 μCi/pmol using T4 polynucleotide kinase (38). Filters were prehybridized and hybridized with the radioactive oligonucleotides at 4°C as previously described (25). The filters were washed with several changes of 5X SET (0.75 M NaCl/0.15 M Tris, pH 8.0/5 mM EDTA) with 0.1% sodium pyrophosphate between 12 and 20°C and were then exposed to film.

Specially Primed cDNA Made from Spleen RNA. The primer was prepared by purifying a 2-kb Hpa II fragment containing the four J_H coding sequences from a subclone derived from the bacteriophage lambda clone ChSp μ27, which contains germline BALB/c DNA. This 2-kb fragment was then digested with the restriction enzymes Dde I, Hae III, Pst I, and Rsa I. This results in a number of restriction fragments, including four that contain part of the J_H coding sequences (21). The restriction fragments were denatured by boiling and annealed to poly(A)$^+$ RNA from spleen. cDNA synthesis primed with the annealed J_H fragments was carried out as described (29). The concentration of α-^{32}P-labeled and unlabeled deoxynucleotide triphosphates was adjusted so that the synthesized material had a specific activity of 2–3 × 10^8 cpm/μg. RNA in the reaction was hydrolyzed with alkali and the single-stranded cDNA was separated from unincorporated nucleotides by gel filtration. The yield of cDNA was ~0.5% the mass of spleen RNA in the reaction; a fourfold stimulation over a reaction with no added primer. Filters were prehybridized at 50°C and hybridized at the same temperature with 5 ng/ml of J_H-primed cDNA for 48 h. Conditions were otherwise as previously described (25). Filters were washed at 50°C in several changes of 5X SET/0.1% sodium pyrophosphate/0.1% SDS before exposure to film.

V Region Probes. The plasmid p107V1 contains the entire gene segment coding for the heavy chain variable region expressed in the S107 myeloma (39). The V_H gene can be separated from the pBR322 vector DNA by digestion with Pst I. The plasmid pVH_5 obtained by the laboratory of Dr. Sam Strober, was provided by Dr. Michael McGrath, Stanford University. This plasmid has a 1-kb Bam HI fragment that contains the heavy chain variable gene expressed by the BCL1 lymphoma (40). The 1-kb Bam HI fragment of pVH_5 and the 445 base pair Pst I fragment of p107V1 were nick translated (41) to a specific activity of 1–8 × 10^8 cpm/μg. Filters hybridized with these probes were handled as described for the J_H-primed cDNA except that the probe was present at a concentration of 0.2–1.0 ng/ml.

Colony Hybridization. Nitrocellulose filters (HATF 13750, Millipore, Bedford, MA) were replica plated and prepared for in situ hybridization as described (42). Duplicate filters were annealed with each probe. For each screening, a positive control filter with colonies containing a heavy chain variable gene segment (MOPC21) was hybridized in parallel (Fig. 1).

Southern Blots. Plasmid DNA was prepared from clones isolated from the cDNA libraries. This DNA was digested with various restriction endonucleases, separated by molecular weight in 1% (wt/vol) agarose gels and transferred to nitrocellulose (43). The filters were then hybridized with the synthetic oligonucleotides or the J_H-primed spleen cDNA as described above.

DNA Sequencing. Restriction fragments were labeled at the 5′ end with ^{32}P-γ-dATP using polynucleotide kinase (37) or labeled at the 3′ end with ^{32}P-α-cordycepin-5′-triphosphate using terminal deoxynucleotidyl transferase (44). The labeled fragments were cut internally with a second enzyme and those isolated fragments were sequenced according to the method of Maxam and Gilbert (37).

Results

cDNA libraries were constructed from three different T cell hybridomas. Some features of the three hybrid cell lines are summarized in Table I. These

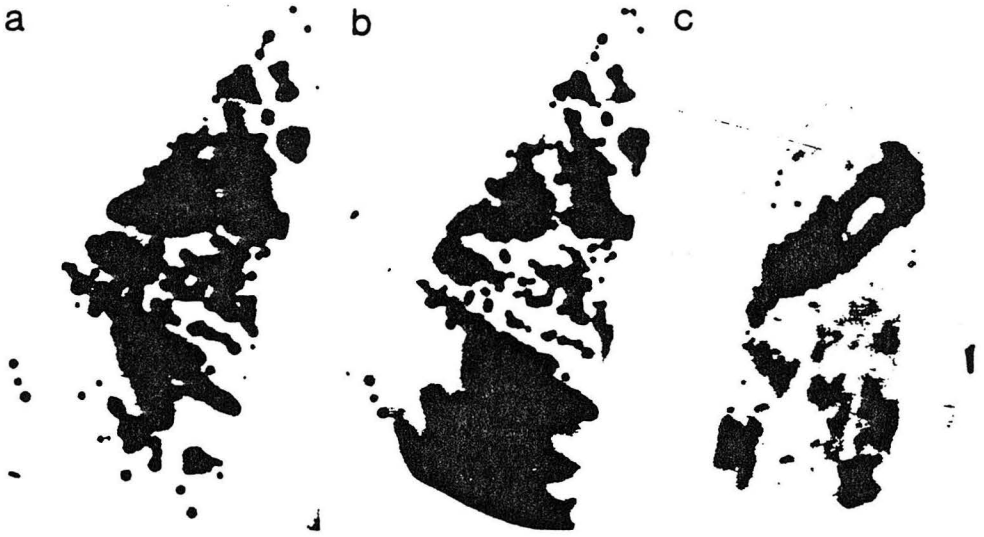


FIGURE 1. Positive control hybridizations to cDNA colonies containing a cloned V_H gene segment. A circular nitrocellulose filter containing DNA from several thousand identical pF9V21 (MOPC21) V_H cDNA clones was cut into sections. Hybridization conditions are described in the Methods section. The hybridization probes were: (*a*) Radiolabeled synthetic oligonucleotides. The filter was exposed to film for 12 h. (*b*) Radiolabeled J_H-primed spleen cDNA. The filter was exposed for 24 h. (*c*) pVH_5 plasmid containing the BCL1 V_H gene. The filter was exposed for 10 h. The same section of filter was hybridized to the synthetic oligonucleotides washed to remove the probe and was then hybridized with the J_H-primed spleen cDNA.

TABLE I
T Cell Hybrids Used to Construct cDNA Libraries

Hybridoma	Normal T cell parent	Fusion partner	Antigen specificity	Functional class	Reference
AODK 10.4	KLH immune B10.D2	AO40.10AG1	KLH + I-A^d	Helper	60
AODH 7.1	HGG immune DBA/2	AO40.10AG1	HGG + I-E^d	Helper	60
395A4.4	GT immune B10.S	BW5147	GT and GAT	Suppressor	59

cells were chosen for several reasons. First, they have retained function that is antigen specific, and in two cases MHC restricted. Second, they grow continuously in the absence of irradiated feeder spleen cells that, if present, could contribute contaminating immunoglobulin sequences to the T cell RNA preparations (21). Third, the hybridomas are specific for structurally unrelated antigens and are therefore likely to employ rather different antigen-binding receptors. If T lymphocytes do transcribe V_H gene segments, this should increase the probability that the probes will have sufficient homology to hybridize with a V_H gene transcript from at least one of the lines. Finally, two hybridomas help B cells secrete antibody and the third secretes a specific suppressor factor. Since most reports of idiotype expression by T lymphocytes involve the helper or suppressor functional subclasses, these types of cells may be best suited for the detection of V_H transcripts (2, 3, 7). To detect virtually any V_H gene transcript present in these cells, we employed two experimental strategies for screening the

cDNA libraries.

Hybridization with Synthetic Oligonucleotides and J_H-Primed cDNA. In the first attempt, two types of probes, synthetic oligonucleotides and a J_H-primed cDNA, were used to screen the libraries. Each probe will hybridize to a variety of V_H gene segments, as well as to some sequences that do not contain V_H genes. The frequency of such non-V_H hybridizing sequences is low, so that a clone hybridizing with both types of probes probably contains a V_H gene segment.

The sequences of the eleven-base synthetic oligonucleotides are 5′ GCA CAG TAA/G TA 3′. These probes are complementary to a highly conserved sequence found at the 3′ end (amino acids 95–98) of mouse heavy chain variable region gene segments. A sequence perfectly complementary to either oligonucleotide is found in 50% (17/34) of the murine V_H genes for which DNA sequence data are available. The degree of homology of the cloned murine V_H gene segments to the oligonucleotides is listed in Table II. Furthermore, in those cases for which no DNA sequence but amino acid sequence data are available, 78% (18/23) of the mouse immunoglobulin variable regions have tyr-tyr-cys-ala at positions 95–98, and therefore may share complete homology to one of the probes (45, 46). However, the oligonucleotide is not long enough to identify unambiguously a V_H gene segment. Given the number of nucleotides of genomic DNA transcribed into RNA, and the random chance of occurrence of an 11-nucleotide sequence, we calculate that any mouse cell, whether it synthesizes immunoglobulin or not, should contain about five species of messenger RNA that will hybridize with each oligonucleotide.[2]

The second type of probe was a cDNA synthesized from spleen RNA (Fig. 2). The spleen contains a relatively high proportion of B cells that should express many different immunoglobulin heavy chains. Therefore, it was possible to use J_H DNA as a primer to stimulate the synthesis of a radioactive single-stranded cDNA complementary to many V_H genes. Using either Southern blots or hybridization to cloned DNA spotted onto nitrocellulose filters, the specifically primed cDNA hybridized to all five V_H gene segments that were tested. These include T15, MOPC21, V_HB2, V14A, and V14B (V14B did not hybridize on one blot [21], but did on a second attempt). Since two of the hybridizing gene segments, T15 and V_HB2, share <60% homology, this probe should hybridize to a large number of different V_H sequences. In addition, the specifically primed cDNA does hybridize to a few cloned DNA segments that do not contain V_H genes (M. Kronenberg, unpublished observations). There are two explanations for hybridization to sequences lacking V_H genes: (*a*) There is a significant amount of cDNA synthesis in the absence of added primer (see Methods) that should not be enriched for immunoglobulin sequences. (*b*) The primer DNA is a mixture of restriction fragments from both the J_H gene segments and the intervening and

[2] A number of measurements have indicated that most mammalian cells, including lymphocytes, contain $\sim 10^4$ different sequences or species of mRNA of average length 2×10^3 nucleotides (48–51). Thus, $\sim 2 \times 10^7$ base pairs ($= 10^4 \times 2 \times 10^3$) of genomic DNA are transcribed into mRNA. The random chance of any one of the four nucleotides occurring at a particular place in a DNA sequence is 1/4; therefore, if we ignore the effects of base composition and nearest neighbors, the probability that an 11-nucleotide sequence will occur is $(1/4)^{11} = 2.5 \times 10^{-7}$. Multiplying the probability of occurrence for the undecamer by the number of nucleotides in the mRNA gives the number of different mRNA species expected to be perfectly complementary to the oligonucleotide, $(2 \times 10^7)(2.5 \times 10^{-7}) = 5$.

TABLE II

Homology of Murine V_H Gene Segments to the Hybridization Probes

Cloned V_H gene segment	Derivation of clone*	Homology of cloned V_H gene segments to the hybridization probes			Comments	Reference
		Synthetic undecamers‡	S107V1§	BCL1§		
V_HT15(V1)	G	11/11	100	56	T15 gene family¶	39, 70
V11	G	11/11	90	59	T15 gene family	70
V13	G	11/11	86	56	T15 gene family	70
S107V1	R	11/11	100	56	T15 gene family	39
M603	R	11/11	98	56	T15 gene family	39
MOPC167	R	10/11	96	56	T15 gene family	71
V14A	G	10/11	75	55	—	S. Crews, unpublished
V14B	G	10/11	73	54	—	S. Crews, unpublished
V_H76	R	10/11	74	55	—	72
V102	G	11/11	58	77	S43 gene family	73
V23	G	10/11	58	78	S43 gene family	73
V3	G	11/11	59	76	S43 gene family	73
V186-1	G	9/11	58	76	S43 gene family	73
V186-2	G	10/11	58	76	S43 gene family	73
S43	R	10/11	56	74	S43 gene family	73
B1-8	R	10/11	58	76	S43 gene family	73
pCH105	G	10/11	60	78	MPC11 gene family	74
pCH108A	G	11/11	59	82	MPC11 gene family	75
pCH108B	G	11/11	57	80	MPC11 gene family	75
MPC11	R	10/11	59	74	MPC11 gene family	76
V_H101	G	10/11	59	57	—	77
V_H101	R	11/11	60	57	Related to V_H101-G	77
PJ14	G	11/11	57	56		78
M141	R	10/11	57	57	Related to PJ14	78
V_H441	G	11/11	71	57		79
UPC10	R	10/11	70	57	Related to V_H441	80
BCL1	R	8/11	56	100	—	40
J558	R	11/11	—**	—**	—	Unpublished
MOPC21	R	11/11	71	58	—	73; this paper
V_HB2	R	11/11	54	74	—	81
V_HB49	R	11/11	56	50	—	81
V_HGAT	R	10/11	57	71	—	25
93G7	R	10/11	56	78	—	82
G5B2.2	R	11/11	58	73	—	83, 84

* *G*, germline V_H gene segment; *R*, rearranged V_H gene segment. Five gene segments considered to be pseudogenes have not been included in this compilation.

‡ Synthetic oligonucleotides we synthesized are: 5′GCACAGTA$^{A}/_{G}$TA3′. The degree of homology is expressed as a fraction, the denominator being the length of the oligonucleotides (11) and the numerator being the maximal number of residues homologous to either undecamer.

§ Percent homology of V_H gene segments to the S107V1 and BCL1 probes. Gaps were introduced where appropriate to compensate for the different lengths of the hypervariable regions. The complete amino acid sequence of the protein produced by the S107 myeloma and the nucleotide sequence of the germline gene (V1) encoding this protein are both available. However, the nucleotide sequence of the rearranged S107V1 cDNA we used as a probe is not complete. Based on the protein sequence we have assumed the rearranged and germline genes are identical, although the possibility of a few silent nucleotide substitutions has not been ruled out.

¶ Gene family denotes a set of closely related sequences. Some members of the T15 family are involved in the response to phosphorylcholine. Some members of the S43 family are involved in the response to NP (4-hydroxy-3-nitrophenylacetyl).

** Insufficient nucleotide sequence data are available.

nearby flanking sequences. Some of the fragments from the noncoding DNA may prime cDNA synthesis from nonimmunoglobulin sequences in spleen RNA. It is possible that some of the nonimmunoglobulin sequences that hybridize with the probe are repeated DNA sequences that are transcribed abundantly in spleen cells. We have not, however, characterized these hybridizing sequences.

To demonstrate that these probes can detect a V_H sequence we screened a cDNA library made with RNA extracted from a B cell hybridoma (25). A colony

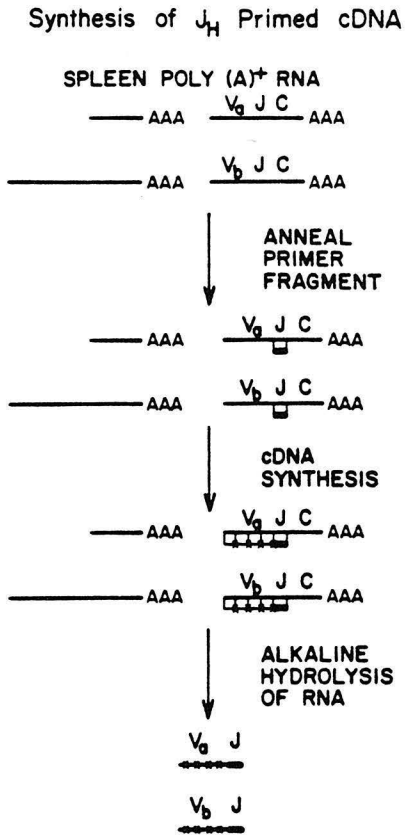


FIGURE 2. Synthesis of J_H-primed cDNA from spleen RNA. Coding sequences on the immunoglobulin heavy chain messenger RNAs are indicated as follows: C, constant region; J, joining gene segment; V_a and V_b, two different V_H gene segments. The J_H primer is indicated by the thick horizontal line. The short vertical lines represent hydrogen-bonded base pairs and the asterisks represent radioactive nucleotides incorporated into cDNA.

that hybridized with both the oligonucleotide and the J_H-primed cDNA (21) was characterized further. The nucleotide sequence of this cDNA clone (Fig. 3) indicates that it encodes the MOPC21 heavy chain variable region synthesized by the P3-X63-Ag8 myeloma parent cell. The clone includes almost the entire V_H gene segment beginning at the codon for amino acid 2 as well as the entire D and J_H4 gene segments.

Having determined that these probes were able to detect B cell V_H gene segments, we screened the three T cell cDNA libraries with the synthetic undecamers. The filters were hybridized and washed under conditions such that 11/11 homology was required to give a positive signal. 54 positive colonies were found (Table III). A single filter containing three hybridizing colonies is shown (Fig. 4). The frequency of positives was low, indicating that these cells do not contain abundant RNA molecules with sequences complementary to the undecamers. Northern blots hybridized with the synthetic oligonucleotide gave a similar result (M. Kronenberg, unpublished observations). In fourteen cases, the colony that hybridized with the oligonucleotide was isolated and the plasmid

```
     Q L V E S G G G L V Q P G G S R K L S C A A S G F
1    TGCAGCTGGTGGAGTCTGGGGGAGGCTTAGTGCAGCCTGGAGGGTCCCGGAAACTCTCCTGTGCAGCCTCTGGATTC  77

     T F S S F G M H W V R Q A P E K G L E W V A Y I S
78   ACTTTCAGTAGCTTTGGAATGCACTGGGTTCGTCAGGCTCCAGAGAAGGGGCTGGAGTGGGTCGCATACATTAGT   152

     S G S S T L H Y A D T V K G R F T I S R D N P K N
153  AGTGGCAGTAGTACCCTCCACTATGCAGACACAGTGAAGGGCCGATTCACCATCTCCAGAGACAATCCCAAGAAC   227

     T L F L Q M T S L R S E D T A M Y Y C A R W G N Y
228  ACCCTGTTCCTGCAAATGACCAGTCTAAGGTCTGAGGACACGGCCATGTATTACTGTGCAAGATGGGGTAACTAC   302
```

FIGURE 3. Nucleotide sequence of a V_H cDNA clone (F9V21) detected with the synthetic oligonucleotides and J_H-primed spleen cDNA. The predicted amino acid sequence is shown above the DNA sequence. The sequence agrees with one previously reported for the MOPC21 V_H gene (73), except for the 209th nucleotide, which was cytidine instead of adenosine. This difference is silent with respect to the predicted amino acid sequence. As noted previously, there are six discrepancies between the published amino acid sequence for MOPC21 and the nucleotide sequence (45, 73). The F9V21 cDNA clone begins in the middle of the codon for the second amino acid, and contains the entire V_H, D, and J_H4 gene segments. A portion of the D and the entire J_H gene segment are not shown. The 11-nucleotide sequence complementary to one of the synthetic nucleotides is underscored.

TABLE III
Colony Hybridization Results

cDNA Library	Synthetic oligonucleotides			J_H-Primed spleen DNA			S107 V_H and BCL1 V_H		
	Colonies hybridized	Positive colonies	Frequency	Colonies hybridized	Positive colonies	Frequency	Colonies hybridized	Positive colonies	Frequency
AODK10.4	140,000	7	1/20,000	26,000	55	1/500	140,000	0	—
AODH7.1	200,000	4	1/50,000	18,000	10	1/1,800	200,000	0	—
395A4.4	920,000	43	1/21,000	200,000	145	1/1,400	920,000	0	—

DNA prepared from the bacterial clone. Southern blots of this plasmid DNA also hybridized with the oligonucleotide, thereby confirming the colony hybridization results (Fig. 5).

Following hybridization with the oligonucleotides, the three cDNA libraries were screened with the J_H-primed cDNA. To test for colonies homologous to both probes, all filters that contained a colony that annealed with the oligonucleotide were hybridized with the cDNA probe. The frequency of positive colonies was 15- to 40-fold higher with the J_H-primed spleen cDNA than was obtained with the synthetic probe (Table III). However, none of the colonies that hybridized with the spleen cDNA also hybridized with the oligonucleotide (Fig. 4). Plasmid DNA isolated from fourteen colonies that hybridized with the undecamer was also tested with this probe. None of the isolated plasmid DNA hybridized with the J_H-primed cDNA (Fig. 5).

Hybridization with Cloned V_H DNA Sequences. Using cloned V_H DNA probes and hybridization conditions of decreased stringency, it is possible to detect cDNA colonies containing V_H genes that are only 55–60% homologous to the probe (S. Crews, unpublished observations, Fig. 1). We therefore screened the three T cell cDNA libraries with V_H probes from the S107 (V_H subgroup III) and BCL1 (V_H subgroup II) tumors. These were chosen because the complete DNA sequences of these V_H gene segments are available and because fragments of the appropriate sizes are easily prepared without contaminating vector or

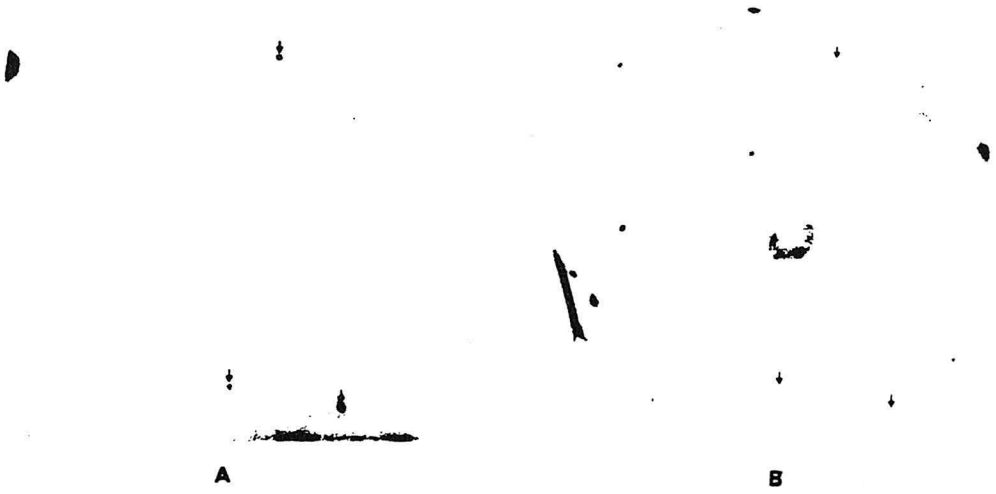


FIGURE 4. Hybridization of the synthetic oligonucleotides and J_H-primed spleen cDNA to T cell cDNA colonies. A single circular filter containing several thousand colonies from the 395A4.4 suppressor T cell cDNA library is shown. (*A*) Hybridization with the synthetic oligonucleotide. The arrows indicate three colonies to which hybridization of the radiolabeled synthetic oligonucleotides was detected in a 72-h exposure. A duplicate filter gave the same pattern of oligonucleotide-positive colonies. (*B*) Hybridization with the J_H-primed spleen cDNA. After incubation of the filter at 50°C to remove the hybridized oligonucleotides, the same filter as shown in (*A*) was hybridized with the radiolabeled J_H-primed spleen cDNA. The signal from positive colonies varies and some nonspecific background is present. The filter was exposed for 48 h. The arrows mark the position of the three oligonucleotide-positive colonies. None of the colonies hybridized with both probes.

much flanking DNA. However, there were no colonies in the three T cell cDNA libraries that hybridized with either of these probes. A positive control filter with colonies containing the MOPC21 V_H gene segment hybridized with pVH_3, which contains the V_H gene expressed in BCL1 (Fig. 1). The MOPC21 and pVH_3 V_H gene segments share only 58% homology. Table II indicates that the V_H gene segments for which nucleotide sequences are available have greater than 56% homology to at least one of our probes. Therefore, it is likely that we could have detected all of these V_H genes.

Discussion

We have utilized several strategies to evaluate whether T cells express any V_H gene segments. To accomplish this, we had to construct DNA probes capable of detecting many different V_H genes. The Northern blot hybridization is the most direct method to test for a particular transcript present even at the level of a single copy per cell (20). For the detection of possible V_H transcripts in T cells, we decided instead to screen large cDNA libraries. There are two reasons for doing this. First, we have found that it is possible to detect V_H sequences <60% homologous to the probe in hybridizations to cDNA colonies (Fig. 1), while >80% homology is required when hybridizing under conditions of moderate stringency to Northern blots (25). This difference may reflect a number of factors, including the concentration of the filter-bound nucleic acid. In addition,

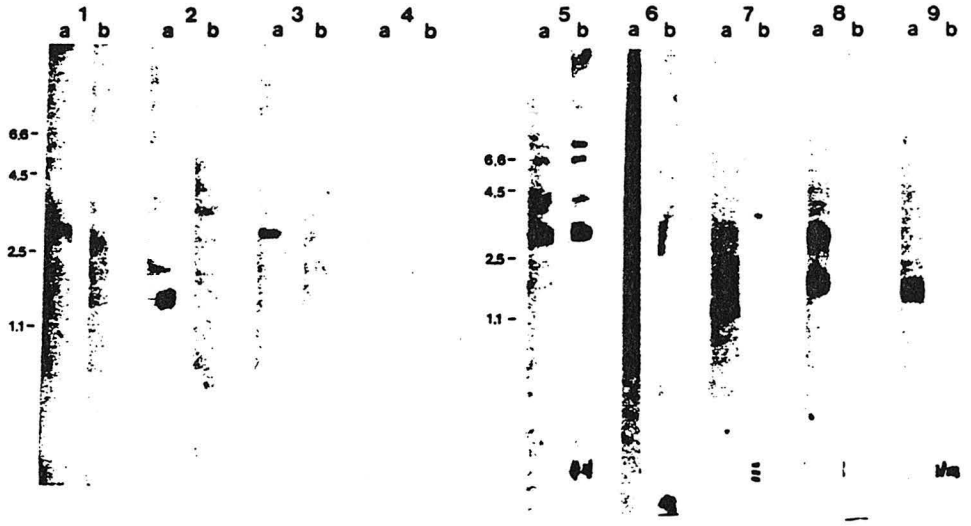


FIGURE 5. Southern blots of some oligonucleotide-positive T cell cDNA colonies. Plasmid DNA was prepared from several T cell cDNA clones that hybridized with the synthetic oligonucleotides. Restriction mapping of the plasmids indicated they contain an average of 800 base pairs of mouse cDNA inserted into the pBR322 cloning vector (4.36 kb). The purified plasmid DNA was digested with the restriction enzymes Eco RI and Pvu II. Digested DNA was electrophoresed on 1% agarose gels and blotted onto nitrocellulose sheets. Migration distances of some molecular weight markers, and their lengths in kilobases, are indicated. Lanes 1–3 and 6–9 contain DNA from separate oligonucleotide-positive colonies isolated from the 395A4.4 library. Lane *4* contains pBR322 vector DNA. Lane *5* contains the MOPC21 V_H cDNA. (*a*) Hybridization with the radiolabeled synthetic oligonucleotides. Exposure was for 3 h. (*b*) Hybridization with the radiolabeled J_H-primed cDNA. The filter was exposed for 24 h.

we have achieved a greater hybridization signal to filter-bound DNA as opposed to RNA, even when both nucleic acids contained identical sequences, were electrophoresed in parallel, and were hybridized to the same probe in the presence of 50% formamide (M. Kronenberg, unpublished observations). Second, the initial strategy to detect a V_H gene segment involved the use of two probes that could hybridize to a number of sequences, including those not containing V_H genes. Since there are multiple sequences present in the poly(A)$^+$ RNA that hybridize to these probes, we detect diffuse smears rather than discrete bands on Northern blots. By contrast, each cloned cDNA in the library is physically separate and could be analyzed individually for the possibly rare sequences that hybridize with both probes.

The cDNA libraries were screened with a synthetic oligonucleotide and a J_H-primed cDNA made from spleen RNA. This method has been characterized extensively by test hybridizations (21) and by the cloning and sequencing of the MOPC21 V_H cDNA from a B cell hybridoma (Fig. 3). Of the $\sim 1.2 \times 10^6$ colonies from three T cell cDNA libraries hybridized with the synthetic oligonucleotide, 54 clones were positive. Because of the relatively short length of the probe, we would have predicted, on statistical grounds, the existence of such colonies even in cDNA made from nonlymphoid RNA. None of these colonies also annealed with the specifically primed spleen cDNA when tested in situ (54

cases) or after Southern blotting the purified plasmid DNA (14 cases). A relatively large number of colonies did, however, hybridize with the spleen cDNA probe. The identification of colonies that hybridized with either one of the two probes provided an internal control and indicates that there was no technical problem with the screening that would have prevented us from identifying clones of interest. Positive control hybridizations with V_H-containing colonies support this conclusion. Since no colonies hybridized with both probes, we conclude that V_H gene segments are absent from these libraries. Hybridization with cloned V_H DNA gene segments under conditions whereby sequences <60% homologous to the probe could be detected constituted a second, independent test for the presence of V_H genes. By hybridizing the cDNA libraries with just two V_H sequences from different heavy chain subgroups, we should be able to detect all of the well-characterized V_H genes. No T cell cDNA colonies hybridized with these probes, confirming the result obtained by the first method.

While B lymphocytes expressing cell-surface IgM contain 100–200 copies of C_μ RNA per cell (17, 47), a messenger RNA for the T cell receptor may not be this prevalent. The bulk of the 5,000–15,000 sequences found in most eucaryotic cells, including lymphocytes, are in the low abundance class (10–20 copies/cell) (48–51). Our calculations indicate that we had a good chance of detecting V_H sequences transcribed at this level (Table IV). For example, if there were a V_H sequence homologous to the probes and present at 10 copies per cell, the probability of detection would be 77% for the AODK 10.4 cDNA library, 88% for the AODH 7.1 cDNA library, and >99% for the 395A4.4 library. The probabilities of detecting a sequence present at 15 or 20 copies per cell are higher. If each library construction and screening were an independent event, then the overall probability of not obtaining a V_H clone that is present at 10 copies per cell becomes extremely low ($0.23 \times 0.12 \times 0.01 = 0.00028$). This analysis cannot exclude the possibility of expression of V_H genes at one or even a few copies per cell. However, there is some indication that these hybrids synthesize a significant amount of receptor protein. The T cell hybridoma AODH 7.1 binds antigen avidly in the presence of the proper antigen-presenting cells

TABLE IV

Probability of Cloning a Rare mRNA from the T Cell cDNA Libraries

Hybridoma	pg RNA per cell (yield)*	pg RNA per cell (total)‡	pg Poly(A)⁺ RNA per cell (yield)*	N (Colonies)§	1 Copy/cell		10 Copies/cell		15 Copies/cell	
					f‖	P**	f	P	f	P
AODK10.4	5.8	9.3	0.15	105,000	1/715,000	0.14	1/71,500	0.77	1/46,500	0.90
AODH7.1	3.3	5.0	0.05	150,000	1/715,000	0.19	1/71,500	0.88	1/46,500	0.96
395A4.4	3.1	4.8	0.13	690,000	1/715,000	0.62	1/71,500	<0.99	1/46,500	<0.99

* RNA yield estimated from absorbance at 260 nm.

‡ Calculated by multiplying the yield of RNA by the recovery of a small amount of ^{3}H-labeled sea urchin RNA added to the preparation. Recovery was at least 60% in each case.

§ N is the number of colonies in the library hybridized; N has been corrected on the assumption that 25% of the RNA is ribosomal or other nonpolyadenylated species. This fraction was estimated following gel electrophoresis of the poly(A)⁺ RNA.

‖ f is the fraction of the mass of total cell poly(A)⁺ RNA present in a given sequence. We assumed each cell contains 0.3 pg of poly(A)⁺ RNA. We considered a sequence of 800 nucleotides, the average length of the cDNA clones, present at either 1, 10, or 15 copies per cell.

** Probability of cloning a gene calculated from the formula. $P = 1 - (1-f)^N$; assumes the probability of detecting a clone is a function of its abundance in the RNA population. See discussion for details.

and almost all the cells in the culture retain this ability (J. Kappler and P. Marrack, unpublished observations). In addition, the 395A4.4 hybridoma constitutively synthesizes both an antigen-binding suppressor factor and an antigen-binding receptor. We therefore consider it unlikely that the receptor mRNA could be present at an extremely low-copy number.

The calculated detection limit depends on a formula that assumes that the probability of detecting a cDNA is solely a function of the abundance of its RNA template in the population (52, 53). This will be true only if the synthesis and cloning of cDNAs uses all templates with equal efficiency. Although factors such as secondary structure of an RNA (54) may influence the ultimate cloning efficiency, there does not appear to be a selection against cloning of heavy chain variable region cDNA. In addition, since the bulk of the first strand of cDNA synthesis is shorter than the average 2-kb length of an mRNA, there is an intrinsic bias towards obtaining cDNA clones containing sequences close to the point of initiation of synthesis. By using a sheared calf thymus DNA primer to initiate cDNA synthesis randomly at many points along the RNA templates, as opposed to an oligo(dT) primer that will initiate synthesis only at the 3′ end, we eliminated any bias towards obtaining clones corresponding to only one end of the messenger RNA. Given these considerations, we feel justified in using the formula shown in Table IV. Finally, in order to calculate the probability of cloning a particular messenger RNA, we needed to estimate the amount of poly(A)$^+$ RNA per cell. Each of the hybrid cells contains between 5 and 10 pg of total RNA (Table IV). We assumed that 0.3 pg of this total is in the poly(A)$^+$ fraction, although our actual yield was substantially lower, between 0.05 and 0.15 pg per cell. Poly(A)$^+$ RNA selected by oligo(dT)-cellulose chromatography will contain a residual poly(A)$^-$ component that is mostly ribosomal RNA. Since this material may give rise to a proportionate fraction of the cDNA colonies in Table IV, we corrected N, the number of colonies screened, to account for the contaminating poly(A)$^-$ RNA.

The calculations are based upon reasonable estimates of the purity of the poly(A)$^+$ RNA and the amount present per cell. However, if we assume there is somewhat more than 0.3 pg of poly(A)$^+$ RNA per cell, or if the poly(A)$^+$ fraction of the RNA preparation is $<75\%$ of the total mass, the basic conclusion remains valid. For example, if the oligo(dT)-passaged RNA were only 50% pure, the probability of detecting a homologous V_H sequence present at 10 copies per cell becomes 63% for the AODK 10.4 cDNA library, 75% for the AODH 7.1 library, and $>99\%$ for the 395A4.4 library.

Although unlikely, it is possible that the three T lymphocytes might express a V_H gene that could not be detected by either of our two screening methods. The DNA sequence homology of mouse V_H gene segments to our probes is presented in Table II. The J_H-primed spleen cDNA hybridized with all of the five V_H gene segments tested (21). In addition, homology with the undecamers does not appear to be restricted to any particular type of heavy chain. Some murine V_H genes coding for proteins in subgroups I, II, and III, as well as three out of four human V_H gene segments (all subgroup III) have complete homology to these probes (55, 56). Even two mouse germline V_λ gene segments have 10/11 matching nucleotides (57, 58). However, the synthetic oligonucleotides will not

hybridize with half of the known murine V_H gene segments. In principle, none of these genes would have been detected by our first method. However, all of the V_H gene segments listed in Table II are >56% homologous to one of the two cloned V_H probes and probably could be detected by the second method.

Estimates of the ability of the probes to detect different V_H gene segments depend upon comparison with the known DNA sequences of relatively few variable gene segments. Almost all of these sequences come from V_H subgroups II and III and a large proportion are members of the gene families involved in binding the NP and PC haptens. It is not certain how well these sequences represent the total germline V_H gene repertoire. To increase the probability of detecting V_H expression, we constructed cDNA libraries from three T cell hybridomas responding to presumably dissimilar antigens. Since each of these cells maintained antigen-specific function, if V_H genes encode the T cell antigen-binding receptor, each should synthesize an RNA containing a V_H gene segment. Because the hybrids are the product of two (395A4.4) or more parental T cells (59, 60), they might be expected to express two or more V_H genes if, as in B cell hybrids, V_H gene segments are transcribed from several chromosome 12 homologues. In addition, it has been suggested that even a single diploid T cell may synthesize more than one V_H RNA (61, 62). Thus, if T cells use the entire V_H gene repertoire, we believe there would have been a V_H segment sufficiently homologous to have been detected by one of our two screening methods. Although it remains formally possible that T cells express selectively a portion of the B cell V_H gene repertoire containing sequences only distantly related to our probes, no such V_H genes have been characterized.

We have presented strong evidence in favor of the proposition that the helper and suppressor T cells tested do not contain RNA with V_H gene segments. This negative conclusion is not, however, completely compelling and two major objections concerning the detection limit and the range of our library screening, have been discussed. A number of unlikely possibilities, including selection against cloning the V_H-containing sequences or a V_H mRNA that is present mostly in the poly(A)$^-$ fraction, have also not been eliminated. However, using different methods, another laboratory has reported that T lymphocytes do not transcribe any V_H gene segments (63).

Many of the genes that have dominated our thinking about the immune response including β_2-microglobulin and the class I and class II products of the MHC, show clear homology to immunoglobulin genes (64–68). The T cell and B cell antigen-binding receptors presumably have somewhat homologous functions. We would be surprised, therefore, if the genes encoding the T cell antigen receptor were to have no homology to immunoglobulin genes. Since multigene families are known to duplicate and diverge (69), it is possible that gene families important for B cell and T cell antigen recognition diverged before or relatively early in vertebrate evolution. Attempts to clone T cell receptor genes using V_H probes might therefore not be feasible, somewhat analogous to attempting to clone V_κ gene segments using V_λ probes. At this point, we can only speculate on the selective forces that may have resulted in separate V gene families expressed in B and T cells.

Summary

We attempted to determine whether T cells express any V_H gene segments. cDNA libraries were constructed from one suppressor and two helper T cell hybridomas. Both the library construction and screening were designed to maximize detection of a wide range of V_H gene segments. One screening method should detect about half of the sequenced V_H genes, while the second should detect most of these genes. The probability of detecting a V_H gene homologous to the probes and present at 10 copies per cell was 77% for one helper cell cDNA library, 88% for the second helper cell library, and >99% for the suppressor cell library. No cDNA clones with V_H gene segments were detected. From this result, we conclude that V_H gene segments are not likely to encode the antigen-specific receptor in the cells we tested.

We thank Drs. Stephen Crews and Roger Perlmutter for helpful discussions, Dr. Suzanna J. Horvath for synthesis of the oligonucleotides, Tim Hunkapiller for development of the data base and computer programs, and Bernita Larsh for help in preparing the manuscript.

Received for publication 21 March 1983.

References

1. Binz, H., and H. Wigzell. 1977. Antigen-binding, idiotypic T-lymphocyte receptors. *Contemp. Top. Immunobiol.* 7:113.
2. Rajewsky, K., and K. Eichmann. 1977. Antigen receptors of T helper cells. *Contemp. Top. Immunobiol.* 7:69.
3. Krammer, P. H. 1981. The T-cell receptor problem. *Curr. Top. Microbiol. Immunol.* 91:179.
4. Binz, H., H. Wigzell, and H. Bazin. 1976. T-cell idiotypes are linked to immunoglobulin heavy chain genes. *Nature (Lond.).* 264:639.
5. Hämmerling, G. J., S. J. Black, C. Berek, K. Eichmann, and K. Rajewsky. 1976. Idiotypic analysis of lymphocytes *in vitro*. II. Genetic control of T-helper cell responsiveness to anti-idiotypic antibody. *J. Exp. Med.* 143:861.
6. Bach, B. A., M. I. Greene, B. Benacerraf, and A. Nisonoff. 1979. Mechanisms of regulation of cell-mediated immunity. IV. Azobenzenearsonate-specific suppressor factor(s) bear cross-reactive idiotypic determinants the expression of which is linked to the heavy-chain allotype linkage group of genes. *J. Exp. Med.* 149:1084.
7. Mozes, E., and J. Haimovich. 1979. Antigen specific T-cell helper factor cross reacts idiotypically with antibodies of the same specificity. *Nature (Lond.).* 278:56.
8. Suzan, M., A. Boned, J. Lieberkind, F. Valsted, and B. Rubin. 1981. The 5936 Ig-idiotype(s): genetic linkage to Ig-C_H loci, T-cell dependence of synthesis and possible specificities. *Scand. J. Immunol.* 14:673.
9. Owen, F. L., A. Finnegan, E. R. Gates, and P. D. Gottlieb. 1979. A mature T lymphocyte subpopulation marker closely linked to the Ig-1 allotype locus. *Eur. J. Immunol.* 9:948.
10. Owen, F. L., R. Riblet, and B. A. Taylor. 1981. The T suppressor cell alloantigen Tsu^d maps near immunoglobulin allotype genes and may be a heavy chain constant region marker on a T cell receptor. *J. Exp. Med.* 153:801.
11. Tokuhisa, T., and M. Taniguchi. 1982. Two distinct allotypic determinants on the antigen-specific suppressor and enhancing T cell factors that are encoded by genes linked to the immunoglobulin heavy chain locus. *J. Exp. Med.* 155:126.
12. Taniguchi, M., T. Tokuhisa, M. Kanno, Y. Yaoita, A. Shimizu, and T. Honjo. 1982.

Reconstitution of antigen-specific suppressor activity with translation products of mRNA. *Nature (Lond.).* 298:172.

13. Tokuhisa, T., and M. Taniguchi. 1982. Constant region determinants on the antigen-binding chain of the suppressor T-cell factor. *Nature (Lond.).* 298:174.
14. Spurll, G. M., and F. L. Owen. 1981. A family of T cell alloantigens linked to Igh-1. *Nature (Lond.).* 293:742.
15. Cory, S., J. M. Adams, and D. J. Kemp. 1980. Somatic rearrangements forming active immunoglobulin μ genes in B and T lymphoid cell lines. *Proc. Natl. Acad. Sci. USA* 77:4943.
16. Forster, A., M. Hobart, H. Hengartner, and T. H. Rabbitts. 1980. An immunoglobulin heavy-chain gene is altered in two T-cell clones. *Nature (Lond.).* 286:897.
17. Kemp, D. J., A. W. Harris, S. Cory, and J. M. Adams. 1980. Expression of the immunoglobulin C_μ gene in mouse T and B lymphoid and myeloid cell lines. *Proc. Natl. Acad. Sci. USA* 77:2876.
18. Kemp, D. J., A. Wilson, A. W. Harris, and K. Shortman. 1980. The immunoglobulin μ constant region gene is expressed in mouse thymocytes. *Nature (Lond.).* 286:168.
19. Kemp, D. J., A. W. Harris, and J. M. Adams. 1980. Transcripts of the immunoglobulin C_μ gene vary in structure and splicing during lymphoid development. *Proc. Natl. Acad. Sci. USA* 77:7400.
20. Kronenberg, M., M. M. Davis, P. W. Early, L. E. Hood, and J. D. Watson. 1980. Helper and killer T cells do not express B cell immunoglobulin joining and constant region gene segments. *J. Exp. Med.* 152:1745.
21. Kronenberg, M., E. Kraig, S. J. Horvath, and L. E. Hood. 1982. Cloned T cells as a tool for molecular geneticists: approaches to cloning genes which encode T-cell antigen receptors. *In* Isolation, Characterization and Utilization of T Lymphocyte Clones. C. Garrison Fathman and Frank Fitch, editors. Academic Press, New York. pp. 467–491.
22. Kurosawa, Y., H. von Boehmer, W. Haas, H. Sakano, A. Trauneker, and S. Tonegawa. 1981. Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. *Nature (Lond.).* 290:566.
23. Zuniga, M. C., P. D'Eustachio, and N. Ruddle. 1982. Immunoglobulin heavy chain gene rearrangement and transcription in murine T cell hybrids and T lymphomas. *Proc. Natl. Acad. Sci. USA* 79:3015.
24. Cerny, J., C. Heusser, W. Reiner, G. J. Hämmerling, and D. D. Eardley. 1982. Immunoglobulin idiotypes expressed by T cells. I. Expression of distinct idiotypes detected by monoclonal antibodies on antigen-specific suppressor T cells. *J. Exp. Med.* 156:719.
25. Kraig, E., M. Kronenberg, J. A. Kapp, C. W. Pierce, L. E. Samelson, R. H. Schwartz, and L. E. Hood. 1983. T and B cells that recognize the same antigen do not transcribe similar heavy chain variable region gene segments. *J. Exp. Med.* 158:192–209.
26. Nakanishi, K., K. Sugimura, Y. Yaoita, K. Maeda, S.-I. Kashiwamura, T. Honjo, and T. Kishimoto. 1982. A T15-idiotype-positive T suppressor hybridoma does not use the T15 V_H gene segment. *Proc. Natl. Acad. Sci. USA* 79:6984.
27. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* 18:5294.
28. Aviv, H., and P. Leder. 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA* 69:1408.
29. Buell, G. N., M. P. Wickens, F. Payvar, and R. T. Schimke. 1978. Synthesis of full length cDNAs from four partially purified oviduct mRNAs. *J. Biol. Chem.* 253:2471.

30. Wickens, M. P., G. N. Buell, and R. T. Schimke. 1978. Synthesis of double-stranded DNA complementary to lysozyme, ovomucoid, and ovalbumin mRNAs. Optimization for full length second strand synthesis by *Escherichia coli* DNA polymerase I. *J. Biol. Chem.* 253:2483.
31. Taylor, J. M., R. Illmensee, and J. Summers. 1976. Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. *Biochim. Biophys. Acta.* 442:324.
32. Villa-Komaroff, L., A. Efstratiadis, S. Broome, P. Lomedico, R. Tizard, S. P. Naber, W. L. Chick, and W. Gilbert. 1978. A bacterial clone synthesizing proinsulin. *Proc. Natl. Acad. Sci. USA* 75:3727.
33. Casadaban, M. J., and S. N. Cohen. 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J. Mol. Biol.* 138:179.
34. Kushner, S. R. 1978. An improved method for transformation of *Escherichia coli* with ColE1 derived plasmids. *In* Genetic Engineering. H. W. Boyer and S. Nicosia, editors. Elsevier/North-Holland Biomedical Press, Amsterdam. pp. 17–23.
35. Matteucci, M. D., and M. H. Caruthers. 1980. The synthesis of oligodeoxypyrimidines on a polymer support. *Tetrahedron Lett.* 21:719.
36. Matteucci, M. D., and M. H. Caruthers. 1980. The use of zinc bromide for removal of dimethoxytrityl ethers from deoxynucleosides. *Tetrahedron Lett.* 21:3243.
37. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499.
38. Houghton, M., A. G. Stewart, S. M. Doel, J. S. Emtage, M. A. W. Eaton, J. C. Smith, T. P. Patel, H. M. Lewis, A. G. Porter, J. R. Birch, T. Cartwright, and N. H. Carey. 1980. The amino-terminal sequence of human fibroblast interferon as deduced from reverse transcripts obtained using synthetic oligonucleotide primers. *Nucleic Acids Res.* 8:1913.
39. Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA : V_H, D and J_H. *Cell.* 19:981.
40. Knapp, M. R., C.-P. Liu, N. Newell, R. B. Ward, P. W. Tucker, S. Strober, and F. Blattner. 1982. Simultaneous expression of immunoglobulin μ and δ heavy chains by a cloned B-cell lymphoma: a single copy of the V_H gene is shared by two adjacent C_H genes. *Proc. Natl. Acad. Sci. USA* 79:2996.
41. Rigby, P. W. J., M. Dieckmann, C. Rhodes, and P. Berg. 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237.
42. Hanahan, D., and M. Meselson. 1980. Plasmid screening at high colony density. *Gene.* 10:63.
43. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503.
44. Tu, C.-P. D., and S. N. Cohen. 1980. 3′-End labeling of DNA with [α-^{32}P] cordycepin-5′ triphosphate. *Gene.* 10:177.
45. Kabat, E. A., T. T. Wu, and H. Bilofsky. 1979. Sequences of immunoglobulin chains. U. S. Department of Health, Education, and Welfare, Bethesda, Maryland. 185 pp.
46. Schilling, J., B. Clevinger, J. M. Davie, and L. Hood. 1980. Amino acid sequence of homogeneous antibodies to dextran and DNA rearrangements in heavy chain V-region gene segments. *Nature (Lond.).* 283:35.
47. Perry, R. P., and D. E. Kelley. 1979. Immunoglobulin messenger RNAs in murine cell lines that have characteristics of immature B lymphocytes. *Cell.* 18:1333.
48. Hastie, N. D., and J. O. Bishop. 1976. The expression of three abundance classes of messenger RNA in mouse tissues. *Cell.* 9:761.
49. Young, B. D., G. D. Birnie, and J. Paul. 1976. Complexity and specificity of polysomal

poly(A$^+$) RNA. *Biochemistry.* 15:2823.
50. Crampton, J., S. Humphries, D. Woods, and R. Williamson. 1980. The isolation of cloned cDNA sequences which are differentially expressed in human lymphocytes and fibroblasts. *Nucleic Acids Res.* 8:6007.
51. Lewin, B. 1980. Eucaryotic chromosomes, Chapt. 24. *In* Gene Expression, Vol. 2. John Wiley & Sons, New York. pp. 694–727.
52. Clarke, L., and J. Carbon. 1976. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire *E. coli* genome. *Cell.* 9:91.
53. Williams, J. G. 1981. The preparation and screening of a cDNA clone bank. *In* Genetic Engineering, Vol. 1. Robert Williamson, editor. Academic Press, New York. pp. 1–59.
54. Payvar, F., and R. T. Schimke. 1979. Methylmercury hydroxide enhancement of translation and transcription of ovalbumin and conalbumin mRNAs. *J. Biol. Chem.* 254:7636.
55. Rabbitts, T. H., D. L. Bentley, W. Dunnick, A. Forster, G. E. A. R. Matthyssens, and C. Milstein. 1981. Immunoglobulin genes undergo multiple sequence rearrangements during differentiation. *Cold Spring Harbor Symp. Quant. Biol.* 45:867.
56. Rechavi, G., B. Bienz, D. Ram, Y. Ben-Neriah, J. B. Cohen, R. Zakut, and D. Givol. 1982. Organization and evolution of immunoglobulin V_H gene subgroups. *Proc. Natl. Acad. Sci. USA* 79:4405.
57. Tonegawa, S., A. M. Maxam, R. Tizard, O. Bernard, and W. Gilbert. 1978. Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. USA* 75:1485.
58. Bernard, O., N. Hozumi, and S. Tonegawa. 1978. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell.* 15:1133.
59. Kapp, J. A., and B. Araneo. 1982. Characterization of antigen-specific suppressor factors from T cell hybridomas. *In* Isolation, Characterization and Utilization of T Lymphocyte Clones. C. Garrison Fathman and Frank Fitch, editors. Academic Press, New York. pp. 137–148.
60. Roehm, N. W., P. Marrack, and J. Kappler. 1982. Antigen-specific, H-2 restricted helper T cell hybridomas. *J. Exp. Med.* 156:191.
61. Janeway, C. A., Jr., H. Wigzell, and H. Binz. 1976. Two different V_H gene products make up the T-cell receptors. *Scand. J. Immunol.* 5:993.
62. Janeway, C., B. Jones, H. Binz, H. Frischknecht, and H. Wigzell. 1980. T-cell receptor idiotypes. *Scand. J. Immunol.* 12:83.
63. Kemp, D. J., J. M. Adams, P. L. Mottram, W. R. Thomas, I. D. Walker, and J. F. A. P. Miller. 1982. A search for messenger RNA molecules bearing immunoglobulin V_H nucleotide sequences in T cells. *J. Exp. Med.* 156:1848.
64. Peterson, P. A., B. A. Cunningham, I. Berggard, and G. M. Edelman. 1972. β_2-Microglobulin—a free immunoglobulin domain. *Proc. Natl. Acad. Sci. USA* 69:1697.
65. Orr, H. T., D. Lancet, R. J. Robb, J. A. Lopez de Castro, and J. L. Strominger. 1979. The heavy chain of human histocompatibility antigen HLA-B7 contains an immunoglobulin-like region. *Nature (Lond.).* 232:266.
66. Steinmetz, M., J. G. Frelinger, D. Fisher, T. Hunkapiller, D. Pereira, S. M. Weissman, H. Uehara, S. Nathenson, and L. Hood. 1981. Three cDNA clones encoding mouse transplantation antigens: homology to immunoglobulin genes. *Cell.* 24:125.
67. Larhammar, D., L. Schenning, K. Gustafsson, K. Wiman, L. Claesson, L. Rask, and P. A. Peterson. 1982. Complete amino acid sequence of an HLA-DR antigen-like β chain as predicted from the nucleotide sequence: similarities with immunoglobulins and HLA-A, -B, and -C antigens. *Proc. Natl. Acad. Sci. USA* 79:3687.
68. Larhammar, D., K. Gustafsson, L. Claesson, P. Bill, K. Wiman, L. Schenning, J.

Sundelin, E. Widmark, P. A. Peterson, and L. Rask. 1982. HLA-DR transplantation antigen α chain is a member of the same protein superfamily as the immunoglobulins. *Cell.* 30:153.

69. Hood, L., J. H. Campbell, and S. C. R. Elgin. 1975. The organization, expression, and evolution of antibody genes and other multigene families. *Annu. Rev. Genet.* 9:305.
70. Crews, S., J. Griffin, H. Huang, K. Calame, and L. Hood. 1981. A single V_H gene segment encodes the immune response to phosphorylcholine: somatic mutation is correlated with the class of antibody. *Cell.* 25:59.
71. Kim, S., M. Davis, E. Sinn, P. Patten, and L. Hood. 1981. Antibody diversity: somatic hypermutation of rearranged V_H genes. *Cell.* 27:573.
72. Bernard, O., and N. M. Gough. 1980. Nucleotide sequence of immunoglobulin heavy chain joining segments between translocated V_H and μ constant region genes. *Proc. Natl. Acad. Sci. USA.* 77:3630.
73. Bothwell, A. L. M., M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, and D. Baltimore. 1981. Heavy chain variable region contribution to the NP^b family of antibodies: somatic mutation evident in a γ2a variable region. *Cell.* 24:625.
74. Cohen, J. B., K. Effron, G. Rechavi, Y. Ben-Neriah, R. Zakut, and D. Givol. 1982. Simple DNA sequences in homologous flanking regions near immunoglobulin V_H genes: a role in gene interaction? *Nucleic Acids Res.* 10:3353.
75. Givol, D., R. Zakut, K. Effron, G. Rechavi, D. Ram, and J. B. Cohen. 1981. Diversity of germ-line immunoglobulin V_H genes. *Nature (Lond.).* 292:426.
76. Zakut, R., J. Cohen, and D. Givol. 1980. Cloning and sequence of the cDNA corresponding to the variable region of immunoglobulin heavy chain MPC11. *Nucleic Acids Res.* 8:3591.
77. Kataoka, T., T. Nikaido, T. Miyata, K. Moriwaki, and T. Honjo. 1982. The nucleotide sequences of rearranged and germline immunoglobulin V_H genes of a mouse myeloma MC101 and evolution of V_H genes in mouse. *J. Biol. Chem.* 257:277.
78. Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa. 1980. Two types of somatic recombination necessary for generation of complete immunoglobulin heavy chain genes. *Nature (Lond.).* 286:676.
79. Ollo, R., C. Auffray, J.-L. Sikorav, and F. Rougeon. 1981. Mouse heavy chain variable regions: nucleotide sequence of a germ-line V_H segment. *Nucleic Acids Res.* 9:4099.
80. Auffray, C., R. Nageotte, B. Chambraud, and F. Rougeon. 1980. Mouse immunoglobulin genes: a bacterial plasmid containing the entire coding sequence for a pre γ2a heavy chain. *Nucleic Acids Res.* 8:1231.
81. Early, P., C. Nottenburg, I. Weissman, and L. Hood. 1982. Immunoglobulin gene rearrangements in normal mouse B cells. *Mol. Cell. Biol.* 2:829.
82. Sims, J., T. H. Rabbitts, P. Estess, C. Slaughter, P. W. Tucker, and J. D. Capra. 1982. Somatic mutation in genes for the variable portion of the immunoglobulin heavy chain. *Science (Wash. DC).* 216:309.
83. Rocca-Serra, J., H. W. Matthes, M. Kaaztinen, C. Milstein, J. Thèze, and M. Fougereau. 1983. Analysis of antibody diversity: V-D-J nucleotide sequence of four anti-GAT antibodies: a pausi-gene system using alternate D-J recombination to generate functionally similar hypervariable regions. *EMBO (Eur. Mol. Biol. Org.) J.* In press.
84. Schiff, C., M. Millili, and M. Fougereau. 1983. Immunoglobulin diversity: analysis of the germline V_H gene repertoire of the murine anti-GAT response. *Nucleic Acids Res.* In press.

APPENDIX II

DIVERSITY AND STRUCTURE OF GENES OF THE α FAMILY OF THE MOUSE T-CELL ANTIGEN RECEPTOR

Published in *Nature*

Diversity and structure of genes of the α family of mouse T-cell antigen receptor

Bernhard Arden, Joan L. Klotz*, Gerald Siu & Leroy E. Hood

Division of Biology, California Institute of Technology, Pasadena, California 91125, USA
* Division of Cytogenetics, City of Hope Medical Center, Duarte, California 91010, USA

We have analysed 19 complementary DNA clones encoding the α-chain of the T-cell antigen receptor derived from thymic transcripts, and find that 15 of them contain partial or complete variable (V_α) genes. Seven of these genes cross-hybridize to over 40 germline V_α gene segments in Southern blot analyses. Of the 19 joining (J_α) sequences examined, 18 seem to be encoded by distinct gene segments, hence the repertoire of J_α gene segments is much larger than those of the immunoglobulin or T-cell receptor β-chain gene families. We suggest that the variable domains of immunoglobulins and T-cell antigen receptors are similar in structure.

THE antigen-specific receptors on T cells, like the immunoglobulin molecules used as antigen-specific receptors on B cells, can recognize and bind an almost unlimited range of antigenic determinants. T-cell antigen recognition differs from immunoglobulin antigen recognition in that it occurs only in the context of another cell-surface protein encoded by the major histocompatibility complex (MHC), a property termed MHC restriction (for review, see ref. 1). This unique property raises the question of whether the structures of T-cell antigen receptors fundamentally differ from those of immunoglobulins.

The T-cell receptor is a disulphide-bridged heterodimer composed of α- and β-chains, each divided into variable (V) and constant (C) regions[2-4]. Studies of the structure and organization of β-chain gene elements have revealed that the V region of the β-chain, like that of the immunoglobulin heavy (H) chain, is encoded by three separate germline gene segments, V_β, diversity (D_β) and joining (J_β), that rearrange during T-cell differentiation to form a functional V_β gene[5,6]. There are two closely linked C_β genes, $C_\beta 1$ and $C_\beta 2$, each associated with a cluster of six functional J_β gene segments[5,7,8]. Each J_β cluster has one D_β gene segment located ~500–700 nucleotides upstream[9,10].

T and B cells use several mechanisms to generate diversity in the antigen-binding V region: (1) germline diversity. Both T and B cells use many different germline *V*, *D* and *J* gene segments to form a *V* gene. (2) Combinatorial joining. Each *D* gene segment seems capable of joining to any downstream *J* gene segment, and any *V* may join to any *D*-*J* combination[11,12]. (3) Somatic mutation. The V_β genes share two somatic mutational mechanisms with their immunoglobulin counterparts. First, there is flexibility in the sites at which the gene segments are joined and this generates diversity in the junctional regions[9,10]. Second, random nucleotides may be added to either side of the *D* gene segments during its joining to the *V* and *J* gene segments[9,10], a mechanism denoted *N*-region diversification[13]. Immunoglobulin V-region genes may undergo somatic hypermutation at a late stage of B-cell development, a phenomenon that has not been seen in most V_β genes[5,11,12,14].

The V_β gene segments differ strikingly from their immunoglobulin κ and H-chain homologues in subfamily organization. A subfamily is defined as those *V* gene segments that are 75% or more similar to one another at the nucleic acid level and hence cross-hybridize in Southern blots under stringent hybridization conditions[15]. The immunoglobulin *V* subfamilies are composed of 4–50 or more members[15-17]. In contrast, the known mouse V_β gene segments represent eight subfamilies, six with only one member, one with two members, and one with three members[11,12].

Recently, mouse cDNA clones of α-chain transcripts have been described[18,19]. The α-chains seem to consist of V, J and C regions. Here we present the sequences of 19 additional α-cDNA clones. Analyses of these 21 α-cDNA sequences reveal that: (1) the V_α gene segments are derived from at least 10 subfamilies each containing 1–10 members; (2) there are many J_α gene segments; (3) the diversification mechanisms for V_α genes are similar to those seen for β-chain and immunoglobulin genes; and (4) the general structure of the α-polypeptides resembles that of immunoglobulin and β-chains in primary and secondary structural characteristics.

Sequences of α-cDNA clones

Nineteen cDNA clones were isolated from a thymic library of inbred C57BL/Ka mice[11] using a C_α probe[20]. The cDNA inserts were subcloned into the M13mp18 bacteriophage vector and the sequences determined using the specific-primer-directed dideoxynucleotide sequencing technique (Fig. 1)[21]. Four of the 19 α-cDNA clones (TA38, TA91, TA45 and TA20) do not have a recognizable V_α gene segment (see below); 17 contain J_α sequences; and 15 contain complete or partial V_α sequences (Figs 1, 2). The α-cDNA sequences in Fig. 1 are divided into leader, *V*, *J* and *C* regions according to their similarities to the V and J segments of immunoglobulins and β-chains. The V_α segment is ~90 residues long and has a leader of 19–22 amino acids (Fig. 1). The N-terminal amino acid of the mature V_α gene product is presumed to be 22 or 23 residues from the first conserved half-cysteine residue according to the general rules that have been deduced for determining the leader peptide cleavage site[22]. The 3′ ends of the V_α sequences have been provisionally designated as the same length as that of a single germline gene segment, $V_\alpha 5H$, a member of the $V_\alpha 1$ subfamily, analysed by Winoto *et al.* in the accompanying paper[20]. The 5′ and 3′ ends of four *J* sequences have been determined by sequence analysis of four germline J_α gene segments (TA84, TA19, TA65 and TA80)[20]. The boundaries of the remaining J_α sequences were provisionally assigned by homology (see Fig. 1 legend). There are nucleotides present between the V_α and J_α sequences that might arise from various different sources (see ref. 20).

Subfamilies of V_α gene segments

The amino-acid sequences of the 15 V_α segments were aligned with one another and with two additional V_α segments from the literature, TT11 (ref. 18) and pHDS58 (ref. 19) (Fig. 2). The V_α segments can be divided into 10 different subfamilies, members of each subfamily being 75% or more homologous with one another at the DNA level. The V_α segments from different subfamilies are 20–54% similar to one another at the protein level (Table 1). This range of homologies is similar to that observed between the V_β and immunoglobulin V_H subfamilies[11]. We have denoted these subfamilies as $V_\alpha 1$ to $V_\alpha 10$ (Fig. 2). Five of these subfamilies ($V_\alpha 1$–5) are represented by two members and five by only one.

To estimate the number of germline V_α gene segments in each subfamily, Southern blot analyses were performed on liver DNA from four different inbred strains of mice using V_α-containing

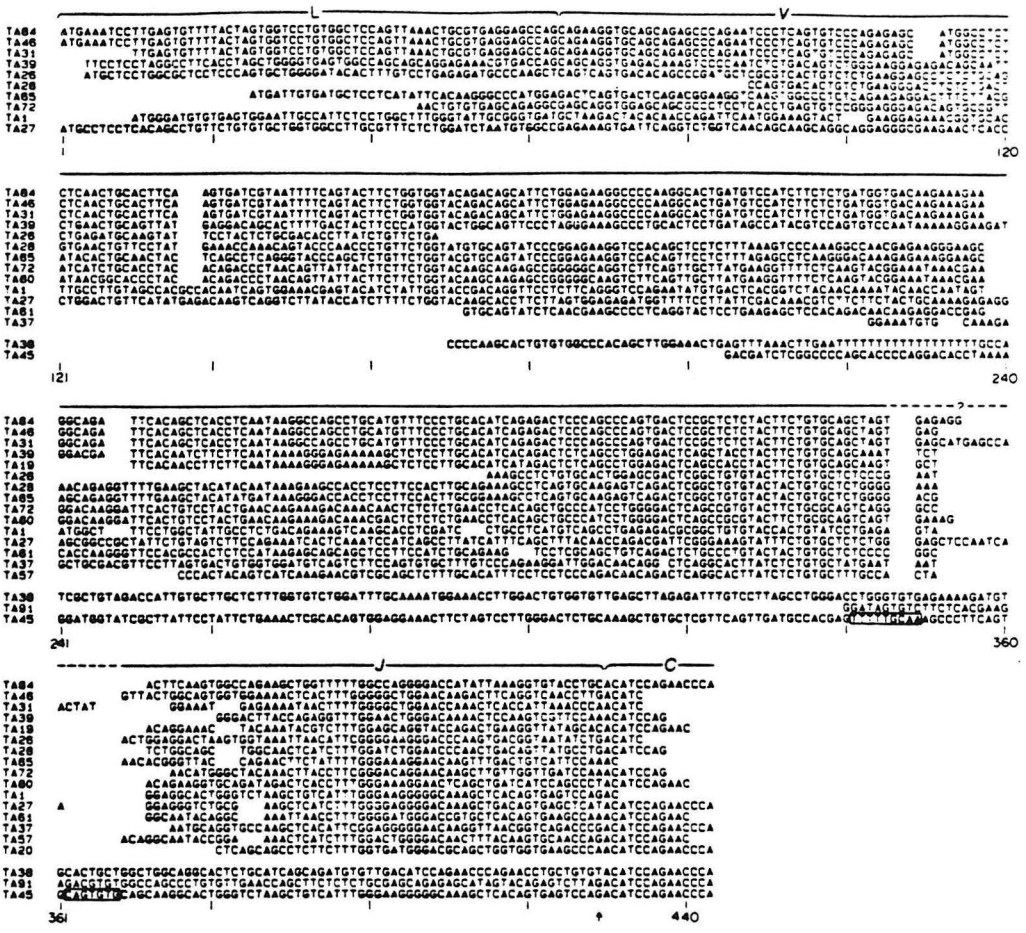


Fig. 1 Nucleotide sequences of 19 α-cDNA clones derived from the thymus. The leader (L), *V*, *J* and *C* regions are indicated. Clones TA38, TA91 and TA45 do not contain V_α sequences. The dotted lines between the *V* and *J* regions indicate junctional diversity arising from any one of several mechanisms (see ref. 20). The 3′ ends of the V_α gene segments were arbitrarily chosen to be identical with the end of the germline V_α5H gene segment[20]. The 5′ ends of the J_α gene segments were determined by comparison with the corresponding germline J_α gene segments (indicated by asterisks)[20] or by arbitrarily assuming that the J_α gene segment begins one codon after the end of the V_α gene segment. Assignments of the 5′ ends of the J_α gene segments for the TA31 and TA27 cDNA clones are complicated by the presence of frameshift mutations at the junctional regions. Gaps are inserted to maximize similarity. The arrow at base 429 indicates diversity at the end of the J_α gene segment (see text). In clone TA45, the recognition sequences for DNA rearrangement are boxed. A 500-bp *Sau*3A1 restriction fragment containing the C_α gene was used as a ^{32}P-labelled, nick-translated probe[37] to screen a λgt10 cDNA library constructed from C57BL/Ka thymus messenger RNA (ref. 38, modified by Barth *et al.*[11]). The cDNA inserts from C_α-containing clones were subcloned into the M13mp18 vector[39]. The subclones were sequenced using the specific-primer-directed dideoxynucleotide chain termination method[21]. Synthetic oligonucleotide probes complementary to both strands of the C_α gene were used to select inserts of both orientations.

Table 1 Similarity of V_α sequences

	TT11	TA84	TA39	TA19	pHDS58	TA26	TA28	TA65	TA72	TA80	TA1	TA27	TA61	TA37	TA57
TT11	—	79	51	56	33	40	31	30	43	42	20	24	38	24	40
TA84	87	—	54	65	31	35	32	31	44	42	22	28	40	22	37
TA39	61	62	—	91	33	38	35	27	44	43	25	33	29	24	43
TA19	66	69	96	—	38	41	34	31	53	56	23	44	35	23	43
pHDS58	47	44	42	46	—	96	52	50	38	37	30	27	46	26	23
TA26	48	42	44	43	96	—	60	56	38	35	35	34	69	25	47
TA28	41	43	45	46	58	63	—	70	36	33	27	26	55	18	33
TA65	42	42	40	44	60	66	83	—	30	29	26	27	54	18	33
TA72	58	60	58	67	49	53	44	44	—	90	25	34	31	16	27
TA80	53	56	56	68	47	49	42	42	95	—	20	33	30	16	30
TA1	37	36	40	41	43	48	41	40	41	40	—	22	24	19	17
TA27	40	42	43	48	40	47	43	42	47	46	37	—	25	21	40
TA61	49	52	46	51	55	69	63	62	50	49	37	39	—	22	41
TA37	36	34	37	39	40	34	28	31	30	32	35	39	29	—	31
TA57	47	41	48	49	42	53	44	41	47	49	40	41	52	37	—

Numbers above the diagonal designate percentage similarity of protein sequences; those below the diagonal compare nucleotide sequences.

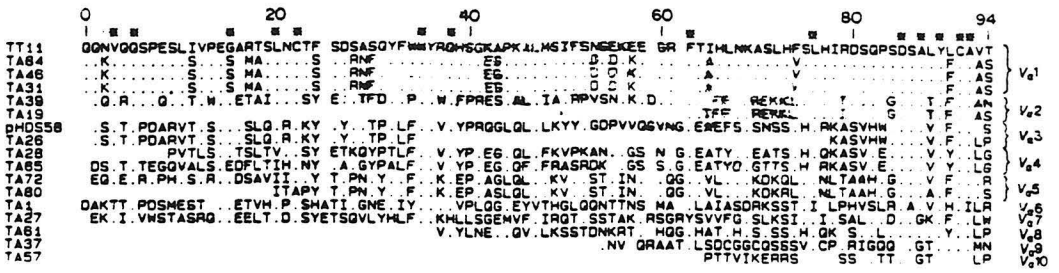


Fig. 2 Protein sequences of 17 V_α gene segments. Clone designations (left) and V_α subfamily designations (right) are given. Dots indicate that the sequence is identical to that of clone TT11. The blanks indicate sequence gaps introduced to maximize homology. The asterisks indicate residues that are relatively conserved in V_H, V_L, V_β and V_κ regions.

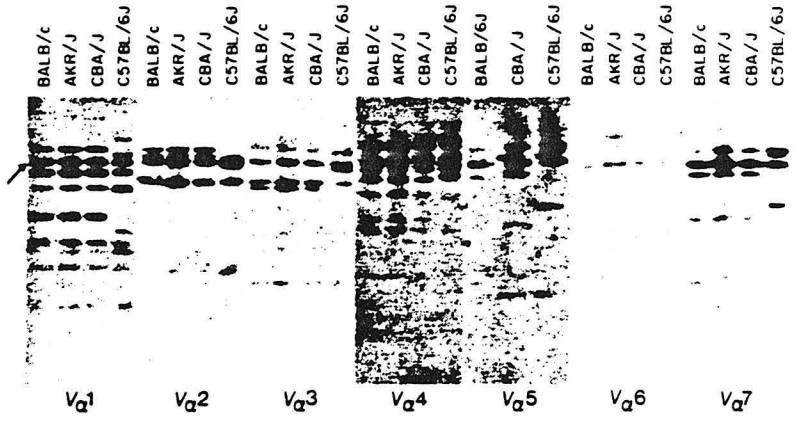


Fig. 3 Southern blot analysis of mouse germline DNA with complete α-cDNA clones from seven V_α subfamilies. Liver DNA from four inbred strains of mice was isolated[40] and *Bam*HI-digested. Similar results were obtained using liver DNA digested with *Eco*RI and *Hin*dIII (not shown). The 9.8-kilobase C_α-hybridizing bands on each filter have been aligned here and are indicated by an arrow. As the DNA migration distances varied slightly from gel to gel, hybridizing bands at the same location on different filters do not necessarily contain DNA fragments of precisely the same size. Each DNA digest (10 μg) was electrophoresed on 0.7% horizontal agarose gels and transferred to nitrocellulose[41]. Hybridizations were for 24 h at 37 °C in 50% formamide, 0.8 M NaCl, 0.1 M Tris *p*H 7.5, 5 × Denhardt's, 100 μg ml^{-1} denatured salmon sperm DNA, 10% dextran sulphate and 0.5 μg ^{32}P-labelled nick-translated cDNA probe. Filters were washed at 68 °C with 1 × SSC containing 0.1% SDS.

*Eco*RI fragments as probes (Fig. 3). Because some of the α-cDNAs contain only a small segment of V_α sequence (Fig. 1), only seven of the V_α subfamilies were analysed. Assuming that each distinct band represents at least one separate V_α gene segment, the V_α subfamilies analysed here range in size from 1 to 10 members, and the minimum number of bands identified for all seven subfamilies is 40. This distribution is similar to that seen in the immunoglobulin κ- and H-chain gene families and contrasts with the V_β gene family, where six of eight subfamilies analysed contained just one member, and the total V_β repertoire is believed to be <21 members[11,12].

Three identical V_α sequences have been observed to be associated with three distinct J_α segments, a surprising result for which we can offer no explanation apart from the possibility that rearrangement and expression of this V_α sequence may be preferred in the thymus. The remaining 12 V_α sequences are all distinct from one another (Fig. 2).

Restriction length polymorphism is observed between inbred C57BL/6J mice and the three other strains of mice analysed (Fig. 3). Differences are seen with all seven V_α probes. In addition, some subfamilies in different mouse strains appear to differ in their V_α gene segment numbers; for example, compare the C57BL/6J strain with the others (Fig. 3). The immunoglobulin H- and κ-chain gene families also exhibit extensive restriction length polymorphism, whereas the V_β gene family shows very limited polymorphism. Variation in band number within particular subfamilies probably represents the duplication and/or deletion of V_α gene segments in the various inbred strains of mice. These data are consistent with the hypothesis that multi-membered *V* subfamilies may undergo gene expansion and contraction by homologous but unequal crossing-over (see ref. 11).

Numerous J_α segments

Of the 17 J_α sequences described in this paper and two from the literature[18,19], 18 different J_α sequences have been identified (Fig. 4). As we found only one repeat, the repertoire of J_α gene segments is probably much larger than 18. We believe that each of these different sequences represents a distinct germline J_α gene segment as they differ considerably from one another in sequence. The J_α segments exhibit a range of similarity at the protein level (32–72%) comparable with that for J_β segments (33–75%) and somewhat lower than that seen for J_κ (55–85%), J_H (57–80%) and J_λ (70–85%) segments. The J_α gene segment repertoire is larger than those of the β-chain (12) or immunoglobulin (4–5) genes. The large repertoire and diversity of the J_α gene segments may have important implications for antigen/MHC recognition, as the J segments in immunoglobulins usually fold to form a contact point of the antigen-binding site.

Nonproductive thymic V_α transcripts

Nonproductive α-cDNAs may arise by joining events which place the *J* gene segment in an improper translational reading frame with respect to the *V* gene segment, by rearrangement of a pseudogene segment or by initiation of transcription 5′ to a germline *J* gene segment or 5′ to an incomplete *D–J* rearrangement. Of the 15 α-cDNAs that have V_α gene segments, 5 seem to be derived from nonproductive transcripts. Three of these may have resulted from the rearrangement of a pseudogene. One clone, TA37, has a frameshift within the V_α gene segment caused by a single nucleotide deletion at position 317 (Fig. 1). The TA80 clone seems to have a base substitution that has changed the cysteine residue at position 22 to alanine (Fig. 2). This cysteine residue is invariant in immunoglobulins and β-chains and is believed to be important for stabilizing the V domain structure. In addition, the V_α and J_α gene segments have joined so that the J_α gene segment is out of the proper translational reading frame. Thus the TA37 and TA80 clones probably represent transcripts derived from the rearrangement of pseudogenes. A third clone, TA26, has a V_α gene segment virtually identical to the previously published pHDS58 V_α gene[19] apart from a large deletion in the centre of the V_α region encompassing 43 residues and creating a translational frame-

shift. Two base changes within the coding region do not lead to changes in amino-acid sequence from the pHDS58 V_α segment and might represent polymorphism between the inbred BALB/c and C57BL/Ka mice. The TA26 V_α gene segment may be a pseudogene that has been rearranged and transcribed (see ref. 23), or alternatively the deletion could represent a cloning artefact. Finally, the TA27 and TA31 clones also contain a V_α-J_α junction which shifts the J_α sequence out of the correct translational reading frame. Nonproductive rearrangements of this type have been reported previously for both immunoglobulin[24] and T-cell receptor β-chain genes (M. Malissen, unpublished data) and are often transcribed[25]. The remaining 10 V_α-containing clones seem to represent productive or functional V_α-J_α rearrangements, although we cannot exclude the possibility that one or more of the partial V_α sequences have mutations in their unsequenced regions that would cause them to be pseudogenes.

α-cDNA clones lacking V_α

Of the 19 cDNA clones analysed, 4 did not contain V_α sequences. One clone, TA20, extends only into the 5′ half of a J_α gene segment. A second, TA45, seems to be either a germline J_α transcript or a transcript from an incomplete D-J rearrangement. Both of these transcription events are known to occur in β-chain genes[9,26]. The TA45 clone contains a J_α sequence and a 5′ recognition sequence for DNA rearrangement, consisting of a conserved heptamer, a spacer sequence of 12 nucleotides and an A/T-rich nonamer. The germline V_α gene segment has the rearrangement recognition sequence with a 23-base-pair (bp) spacer[20]. In immunoglobulin and β-chain genes, gene-segment rearrangements always occur between recognition sequences with 12- (one turn of the DNA helix) and 23-bp (two turns) spacers[5,6,27,28]. Hence, this one-turn recognition sequence can join to the two-turn recognition sequences of the V_α genes[20]. As the germline sequence of this J_α gene segment has not been determined, we do not know whether the recognition sequence lies to the 5′ side of a D_α or a J_α gene segment. Finally, two of the cDNA clones encode complete C_α, but no J_α or V_α regions. The sequences to the 5′ side of the C_α genes are not similar to each other or to the intron sequences flanking the germline C_α gene (A. Winoto, unpublished data). Presumably, these cDNA clones represent aberrant RNA splicing events.

V_α gene diversification mechanisms

The α-chain genes apply three mechanisms used by the immunoglobulin and β-chain genes for diversification.

Germline. Southern blot analyses with V_α probes and the V_α sequence analyses suggest that at least 10 different V_α subfamilies exist which encode >40 V_α gene segments. Only one V_α sequence is repeated in the sample of cDNAs analysed. A comparable analysis of 22 V_β sequences revealed 11 repeat V_β sequences and only 13 distinct V_β gene segments[11]. Statistical analyses of thymus β-chain cDNA clones indicate that there are ≤21 germline V_β gene segments[11] and perhaps only two D_β gene segments[11,14]. Hence the diversity of the V_α gene segment repertoire seems to be considerably larger than that of the V_β gene segments. The germline J_α repertoire seems to be quite large in that 18 out of 19 sequences sampled are different. This J_α diversity greatly increases the germline diversity at the carboxy-terminal end of the V_α region. Although provisional alignments of the different α-cDNA clones with the available germline V_α and J_α sequences reveal additional nucleotides at the V-J junctions (Fig. 1), additional germline α-sequences will have to be determined before the existence of D_α gene segments can be established (see ref. 20).

Combinatorial joining. Two examples of combinatorial joining in the α-family have been observed in the sample of α-cDNA sequences examined. First, the same V_α sequence in clones TA84, 46 and 31 has been joined to three distinct J_α gene segments (Fig. 1). Similarly, the same J_α sequence has been joined to two distinct V_α gene segments (TA28 and A. Winoto, unpublished data). Therefore, the combinatorial joining of V_α and J_α gene segments can occur in the α-family.

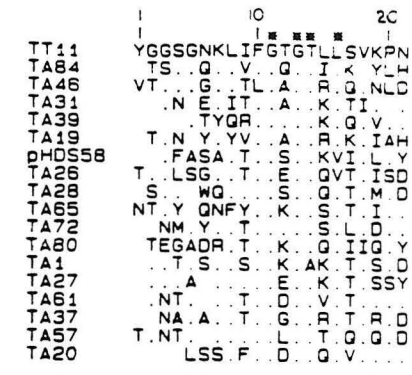


Fig. 4 Protein sequences of J_α segments. The N- and C-terminal boundaries of these sequences have been approximated using the germline sequences of six J_α gene segments[20]. Asterisks indicate amino acids conserved in the J_α, J_β, J_H and J_L regions. Dots indicate identity to the TT11 sequence. Gaps are inserted to maximize similarity. The boundaries of the remaining J_α regions were provisionally assigned to the second residue following the putative C-terminal end encoded by the V segment. Diversity in the C-terminal residue is generated by RNA splicing (see text).

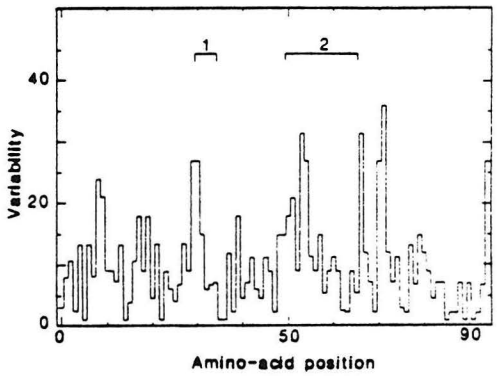


Fig. 5 Variability plot of V_α regions. Variability was determined according to the method of Wu and Kabat[31]. Nine V_α segments were analysed. The variability at each amino-acid position N is calculated as: Variability$_N$ = (No. of different amino acids that occur at N)/(Frequency of most common amino acid at N). For comparison, the first and second hypervariable regions of V_H segments are indicated.

Somatic mutation. The delineation of junctional and N-region diversification mechanisms requires a comparison between germline and rearranged V_α sequences. These mechanisms are used in the V_α gene family, as demonstrated in the accompanying paper[20]. The fact that three V_α gene segments are identical to one another suggests that somatic hypermutation has not occurred in these α-chain genes. However, somatic hypermutation occurs late in B-cell development. Here we are analysing antigen-receptor genes at an early stage of T-lymphocyte development, possibly before somatic hypermutation. Data on the receptors of mature T cells will be necessary to determine whether somatic mutation does occur in V_α genes.

Diversification of a special type occurs at the J_α-C_α boundary. The last nucleotide in the J_α gene segments may be any of the four bases (Figs 1, 4). In the immunoglobulin and β-chain gene families the final nucleotide of the J gene segment is invariant. As RNA splicing joins the last nucleotide of the J gene segment to the first two nucleotides of the C gene, four different junctional codons are generated at the J_α-C_α junction: asparagine (AAC), histidine (CAC), aspartic acid (GAC), and tyrosine (TAC). It will be interesting to determine whether this boundary

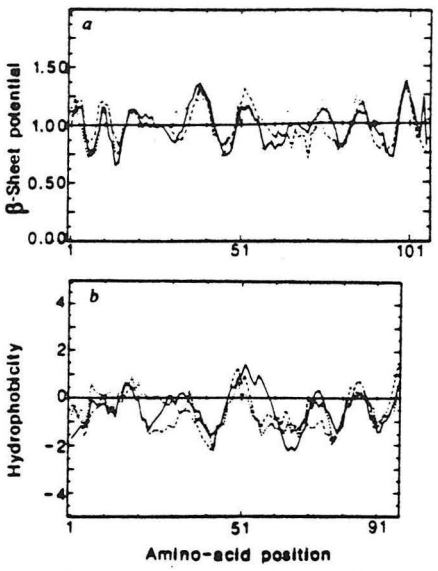


Fig. 6 Secondary structural analyses of V_α (solid lines), V_H (dotted lines) and V_β (dashed lines) regions. *a*, Plot indicating the β-pleated sheet-forming potential of V_β, V_H and V_α segments using the algorithm of Chou and Fasman[34,35]. *b*, Plot of the relative hydrophobicity of the side chains of the amino acids constituting the V_α, V_β and V_H regions, using the algorithm of Kyte and Doolittle[33]. The V_β sequences and V_H sequence compilations are from Barth *et al.*[11].

diversity has any functional significance for the T-cell antigen-binding receptor.

Structure of the V_α region

We have analysed the primary and secondary structures of the V_α regions to determine whether they resemble the V regions of β-chains and immunoglobulins.

Both the V_α and V_β regions share many conserved residues with both the V_H and variable light-chain (V_L) regions of immunoglobulins, including 14 residues in the V and 4 residues in the J regions (marked by asterisks in Figs 2, 4). An analysis of the three-dimensional structures of the V domains in several immunoglobulin molecules demonstrates that there are many highly conserved residues important for stabilization of V_L–V_H interactions and for intrachain V_L or V_H interactions[29,30]. Most of these residues are conserved in the V_α and V_β regions.

Another approach to elucidate primary structural patterns in the T-cell receptor and immunoglobulin V regions is the variability plot of Wu and Kabat[31]; this analysis examines the distribution of variation of each residue position between members of a set of similar sequences. In immunoglobulin and β-chain V regions, two regions of relative hypervariability are noted, with a third hypervariability[31] region positioned at the junction of the *V*, *D* and *J* gene segments[11,32]. A variability plot of the V_α sequences does not differ dramatically from the V_H and V_β variability plots (Fig. 5 and refs 11, 12). With the limited number of V_α sequences analysed, one must be cautious in drawing generalizations for the overall variability in this set is high. However, the regions corresponding to the two classically defined hypervariable regions (residues 30–35 and 51–60) seem to have increased variability. We note two additional areas of hypervariability (residues 8–9 and 66–71) but feel their existence is uncertain until more V_α-sequences are available for analysis. The junction of the V_α and J_α segments generates an additional hypervariable region (not shown).

Two further analyses have been conducted on the V_α sequences to assess the nature of their secondary structure. The Kyte-Doolittle hydrophobicity profile[33] plots the distribution of the relative hydrophobicities of the amino-acid side chains of a particular protein. The Chou-Fasman method[34,35] determines the relative potential for β-pleated sheet formation. Both these analyses measure properties that are thought to reflect important secondary structural features of the protein. The Kyte-Doolittle hydrophobicity and the Chou-Fasman plots of the V_α, V_β, V_H and V_κ (not shown) sequences are very similar (Fig. 6). The plots represent averages of multiple sequences to minimize unusual individual variations. These data indicate that the secondary structures of the V_α, V_β, V_H and V_κ regions are similar to each other.

Thus, several lines of evidence suggest that general primary and secondary structural features are conserved in the V_α, V_β, V_H and V_κ regions: (1) they share highly conserved amino-acid residues, most of which are important for stabilization of the highly conserved immunoglobulin fold; (2) the variability plots demonstrate variability consistent with three hypervariable regions; and (3) the hydrophobicity and Chou-Fasman plots suggest similar potential for β-pleated sheet formation. These data suggest that the T-cell antigen receptor and immunoglobulin molecules have similar structure. Moreover, as immunoglobulins may also be MHC-restricted[36], it is unnecessary to postulate any special sites, apart from the classical antigen-binding site, for binding properties of T-cell MHC-restricted antigen receptors.

We thank Debbie Maloney and Marty Garcia for technical assistance, Marilyn Tomich and Dr Suzanna Horvath for the synthetic oligonucleotide used in the DNA sequencing, Drs Howard Gershenfeld, Craig Okada and Irving Weissman for the thymocyte cDNA library, Drs Richard Barth, Byung Kim, Joan Kobori, Mitchell Kronenberg, Ulf Landegren, Nilabh Shastri, Martha Zuniga, Tim Hunkapiller and Astar Winoto for critical comments on the manuscript, and Bertha Jones, Gwen Anastasi and Susan Mangrum for help in preparing the manuscript. This work was supported by grants from the NIH, T Cell Sciences Inc., the State of California Department of Health Services (to J.L.K.) and the Deutsche Forschungsgemeinschaft (to B.A.).

Received 10 May; accepted 2 July 1985.

1. Golub, E. S. *Cell* **21**, 603–604 (1980).
2. Allison, J. P., McIntyre, B. W. & Bloch, D. *J. Immun.* **129**, 2293–2300 (1982).
3. Haskins, K. *et al. J. exp. Med.* **157**, 1149–1169 (1983).
4. Meuer, S. C. *et al. J. exp. Med.* **157**, 705–719 (1983).
5. Chien, Y., Gascoigne, N. R. J., Kavaler, J., Lee, N. E. & Davis, M. M. *Nature* **309**, 322–326 (1984).
6. Siu, G. *et al. Cell* **37**, 393–401 (1984).
7. Malissen, M. *et al. Cell* **37**, 1101–1110 (1984).
8. Gascoigne, N. R. J., Chien, Y., Becker, D. M., Kavaler, J. & Davis, M. M. *Nature* **310**, 387–391 (1984).
9. Siu, G. *et al. Nature* **311**, 344–350 (1984).
10. Kavaler, J., Davis, M. M. & Chien, Y. *Nature* **310**, 421–423 (1984).
11. Barth, R. K. *et al. Nature* **316**, 517–523 (1985).
12. Patten, P. *et al. Nature* **312**, 40–46 (1984).
13. Alt, F. W. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4118–4122 (1982).
14. Goverman, J. *et al. Cell* **40**, 859–867 (1985).
15. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. *Cell* **25**, 59–66 (1981).
16. Brodeur, P. H. & Riblet, R. *Eur. J. Immun.* **14**, 922–930 (1984).
17. Cory, S., Tyler, B. M. & Adams, J. M. *J. molec. appl. Genet.* **1**, 103–116 (1981).
18. Chien, Y. *et al. Nature* **312**, 31–35 (1984).
19. Saito, H. *et al. Nature* **312**, 36–40 (1984).
20. Winoto, A., Mjolsness, S. & Hood, L. *Nature* **316**, 832–836 (1985).
21. Strauss, E., Kobori, J., Siu, G. & Hood, L. *Analyt. Biochem.* (submitted).
22. von Heijne, G. *Eur. J. Biochem.* **133**, 17–21 (1983).
23. Takahashi, N., Noma, T. & Honjo, T. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5194–5198 (1984).
24. Coleclough, C., Perry, R. P., Karjalainen, K. & Weigert, M. *Nature* **290**, 372–378 (1981).
25. Kwan, S., Max, E., Seidman, J. G., Leder, G. & Scharff, M. D. *Cell* **26**, 57–66 (1981).
26. Yoshikai, Y. *et al. Nature* **312**, 521–524 (1984).
27. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. *Cell* **19**, 981–992 (1980).
28. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286**, 676–683 (1980).
29. Segal, D. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **71**, 4298–4302 (1974).
30. Saul, F. A., Amzel, L. M. & Poljak, R. J. *J. biol. Chem.* **253**, 585–597 (1978).
31. Wu, T. T. & Kabat, E. A. *J. exp. Med.* **132**, 211–250 (1970).
32. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. *Sequences of Proteins of Immunological Interest* (US Department of Health and Human Services, Washington, DC, 1983).
33. Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105–132 (1982).
34. Chou, P. Y. & Fasman, G. D. *Biochemistry* **13**, 211–222 (1974).
35. Chou, P. Y. & Fasman, G. D. *Biochemistry* **13**, 222–245 (1974).
36. Wylie, D. E., Sherman, L. A. & Klinman, N. R. *J. exp. Med.* **155**, 403–414 (1982).
37. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
38. Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning Techniques: A Practical Approach* (ed. Glover, D. M.) (IRL, Oxford, 1984).
39. Yanish-Perron, C., Vieira, J. & Messing, J. *Gene* **33**, 103–119 (1985).
40. Blin, N. & Stafford, D. W. *Nucleic Acids Res.* **3**, 2303–2308 (1976).
41. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).